APPENDIX A

DEFINITIONS OF TERMS RELEVANT TO PRA AND REFERENCES FOR FURTHER READING

DEFINITIONS OF TERMS

This guidance adopts the definitions of variability, uncertainty, and Monte Carlo simulation found in EPA's *Guiding Principles for Monte Carlo Analysis* (1997a). Definitions for the specialized terms pertaining to probabilistic analysis are presented in this appendix. Note that if a definition uses a term that is defined elsewhere in the same list of definitions, it is highlighted in bold text. Definitions are also presented at the beginning of each chapter, sometimes with additional terms and examples that are relevant to concepts presented in the chapter.

Definitions of Terms Used in PRA

50th percentile	The number in a distribution such that half the values in the distribution are greater than the number and half the values are less. The 50th **percentile** is equivalent to the **median**.
95th percentile	The number in a distribution such that 95% of the values in the distribution are less than or equal to the number and 5% are greater.
95 % Upper Confidence Limit for a Mean	The 95 percent upper confidence limit (95% UCL) for a mean is defined as a value that, when repeatedly calculated for randomly drawn subsets of size *n*, equals or exceeds the true population mean 95 percent of the time. Although the 95% UCL provides a conservative estimate of the mean, it should not be confused with a 95th **percentile**. As sample size increases, the difference between the UCL for the mean and the true mean decreases, while the 95th percentile of the distribution remains relatively unchanged, at the upper end of the distribution. EPA's Superfund program has traditionally used the 95% UCL for the mean as the concentration term in **point estimates** of **RME** for human health risk assessment (U.S. EPA, 1992; 1997b).
ARARs	Applicable or relevant and appropriate requirements. The NCP states that ARARs shall be considered in determining remediation goals. If an ARAR meets the requirements of the NCP (U.S. EPA, 1990) for protectiveness, it may be selected as a site-specific cleanup level.
Assessment Endpoint	A term usually associated with ecological risk assessment; a specific expression of the population or ecosystem value that is to be protected. It can be characterized both qualitatively and quantitatively in the risk assessment.
Backcalculation	A method of calculating a PRG that involves algebraic rearrangement of the risk equation to solve for concentration as a function of risk.

Definitions of Terms Used in PRA

Background Exposure	Exposures that are not related to the site. For example, exposure to chemicals at a different time or from locations other than the exposure unit of concern. Background sources may be either naturally occurring or anthropogenic (man-made).
Bayesian Analysis	Statistical analysis that describes the probability of an event as the degree of belief or confidence that a person has, given some state of knowledge, that the event will occur. Bayesian **Monte Carlo** combines a prior probability distribution and a **likelihood function** to yield a posterior distribution (see Appendix E for examples). Also called subjective view of probability, in contrast to the frequentist view of probability.
Bootstrap	A method of sampling actual data at random, with replacement, to derive an estimate of a population **parameter** such as the arithmetic **mean** or the standard error of the mean. The sample size of each bootstrap sample is equal to the sample size of the original data set. Both parametric and nonparametric bootstrap methods have been developed.
Boxplot	Graphical representation showing the center and spread of a distribution, sometimes with a display of outliers (e.g., Figure 4-3). This guidance uses boxplots to represent the following **percentiles**: 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th}.
Cancer Slope Factor (CSF)	A plausible upper-bound estimate of the probability of a response per unit dose of a chemical over a lifetime. The CSF is used to estimate an upper-bound probability of an individual developing cancer as a result of a lifetime of exposure to a particular level of a potential carcinogen.
C-term	The concentration variable used in exposure assessment. Concentration terms are expressed in units applicable to the media of concern (e.g., mg/L for water, :g/m^3 for air; mg/kg for soil and dust.
Central Limit Theorem	If random samples of size *n* are repeatedly drawn from a population of any distribution, the distribution of sample means converges to the normal distribution. The approximation improves as *n* becomes large.
Central Tendency Exposure (CTE)	A risk descriptor representing the average or typical individual in the population, usually considered to be the arithmetic **mean** or **median** of the risk distribution.
Cleanup Level	A chemical concentration chosen by the risk manager after considering both RGs and the nine selection-of-remedy criteria of the NCP (U.S. EPA, 1990; 40CFR 300.430(e)(9)(iii)). Also referred to as Final Remediation Levels (U.S. EPA, 1991), chemical-specific cleanup levels are documented in the Record of Decision (ROD). A cleanup level may differ from a **PRG** because risk managers may consider various uncertainties in the risk estimate, the technical feasibility of achieving the PRG, and the nine criteria outlined in the NCP.
Coefficient of Variation	Ratio of the standard deviation (SD) to the arithmetic **mean** (AM) (CV = **SD/AM**). Dimensionless measure of the spread of a distribution, therefore, useful for comparing **PDFs** for different random variables.

Definitions of Terms Used in PRA

Confidence Interval	An interval characterized by upper and/or lower estimates of an unknown quantity. The confidence level is the probability that the confidence interval contains the true value. Confidence intervals can be determined for any **parameter** of a probability distribution (e.g., arithmetic mean, 95th percentile).
Continuous Variable	A **random variable** that can assume any value within an interval of real numbers (e.g., concentration).
Correlation	A quantitative relationship between two or more input variables of a model (e.g., body weight, inhalation rate, skin surface area). In analyses involving time-dependent variables, a change in one variable is accompanied by a change in another time-dependent, correlated variable. Ignoring correlations in **PRA** may lead to unrealistic combinations of values in a risk calculation. Correlations can also be defined as relationships between inputs and outputs.
Cumulative Distribution Function (CDF)	A representation, generally a function or graph (e.g., Fig. 1-1) of the cumulative probability of occurrence for a **random** independent variable. The CDF is obtained from the **PDF** by integration in the case of a continuous random variable and by summation for **discrete random variables**. Each value *c* of the function is the probability that a random observation *x* will be less than or equal to *c*.
Discrete Variable	A **random variable** that can assume any value within a finite set of values (e.g., number of rainfall events in one month) or at most a countably infinite set of values.
Empirical Distribution	A distribution obtained from actual data and possibly smoothed with interpolation techniques. Data are not fit to a particular parametric distribution (e.g., normal, lognormal), but are described by the percentile values.
Expert Judgment	An inferential opinion of a specialist or group of specialists within an area of their expertise. Experts judgment may be based on an assessment of data, assumptions, criteria, models, and **parameters** in response to questions posed in the relevant area of expertise (see Chapter 3 [Section 3.8], and Appendix E).
Exposure Assessment	The qualitative or quantitative estimate (or measurement) of the magnitude, frequency, duration, and route of exposure. A process that integrates information on chemical fate and transport, environmental measurements, human behavior, and human physiology to estimate the average doses of chemicals received by individual receptors. For simplicity in this guidance, exposure encompasses concepts of absorbed dose (i.e., uptake and bioavailability).
Exposure Point Concentration	The contaminant concentration within an exposure unit to which receptors are exposed. Estimates of the EPC represent the concentration term used in exposure assessment.
Exposure Unit	A geographic area where exposures occur to the receptor of concern during the time of interest. Receptors may be human or ecological (e.g., plants, birds, fish, mammals). For purposes of **PRA**, probability distributions for exposure and toxicity variables apply equally to all members of a population at a given exposure unit. Ecological exposure units often consider habitat and seasonality factors that enhance exposure in a spatial area usually related to home ranges.

Definitions of Terms Used in PRA

Frequency Distribution	A graph or plot that shows the number of observations that occur within a given interval; usually presented as a histogram showing the relative probabilities for each value. It conveys the range of values and the count (or proportion of the sample) that was observed across that range (see Figure 1-1, Figure 4-1).
Geometric Mean (GM)	The n^{th} root of the product of n observations. For lognormal distributions, the GM is equal to the median and is less than the arithmetic **mean** (AM). For normal distributions, all three measures of central tendency (GM, AM, **median**) are equal.
Geostatistics	Branch of statistics that focuses on data that have a spatial or geographic components (e.g., chemical concentrations in soil or groundwater).
Goodness-of-Fit (GoF) Test	A method for examining how well (or poorly) a sample of data can be described by a hypothesized probability distribution for the population. Generally involves an **hypothesis test** in which the null hypothesis H_0 is that a random variable X follows a specific probability distribution F_0. That is, H_0: $F = F_0$ and H_a: F ... F_0.
Hypothesis Testing	Statistical test of an assumption about a characteristic of a population. The goal of the statistical inference is to decide which of two complementary hypotheses is likely to be true.
Indepedence	Two events A and B are independent if knowing whether or not A occurs does not change the probability that B occurs. Two random variables X and Y are independent if the joint probability distribution of X and Y factors into the product of the individual marginal probability distributions. That is, f(X, Y) = f(X) ∧ f(Y). Independence of X and Y is *not* synonymous with zero correlation (i.e., Cor(X, Y) = 0). If X and Y are independent, then Cor(X, Y) = 0; however, the converse is not necessarily true (Law and Kelton, 1991) - X and Y may be related in a nonlinear fashion but still maintain Cor(X, Y) = 0.
Independent and Identically Distributed (IID)	Random variables that are **independent** and have the same probability distribution of occurrence.
Individual-Level Effect	An assessment endpoint that focuses on protecting a hypothetical or real individual in a population. Individual-based models may account for unique exposure and toxicological response to chemicals among individual receptors.
Iterative Truncation	A method of calculating a **PRG** that involves developing an expression for the concentration term in which high-end values are "truncated" to reduce the maximum concentration, and calculating risks associated with the reduced concentration. The method may be repeated with consecutively lower truncation limits until risk is acceptable. Iterative truncation methods avoid difficulties associated with applying **Monte Carlo analysis** to a **backcalculation**.
Kriging	A statistical interpolation method that selects the best linear unbiased estimate of the parameter in question. Often used as a **geostatistical** method of spatial statistics for predicting values at unobserved locations based on data from the surrounding area. Information on fate and transport of chemicals within the area lacking data can be incorporated into kriged estimates.

Definitions of Terms Used in PRA

Kurtosis	The measure of peakedness of a distribution. A uniform distribution has a lower kurtosis than a peaked distribution such as the normal and lognormal distribution. Kurtosis is referred to as the 4th central **moment of a distribution**.
Latin Hypercube Sampling (LHS)	A variant of the **Monte Carlo sampling** method that ensures selection of equal numbers of values from all segments of the distribution. LHS divides the distribution into regions of equal sampling coverage. Hence, the values obtained will be forced to cover the entire distribution. It is more efficient than simple random sampling, i.e., it requires fewer iterations to generate the distribution sufficiently.
Likelihood Function	A term from **Bayesian** statistics referring to a probability distribution that expresses the probability of observing new information given that a particular belief is true.
Mean	Arithmetic mean (AM) or average; the sum of all observations divided by the number of observations. Referred to as the first central **moment of a distribution**.
Microexposure Event (MEE) Analysis	A method of assessing risk based on an aggregate sum of a receptor's contact with a contaminated medium. MEE analysis simulates lifetime exposure as the sum of many short-term, or "micro" exposures (see Appendix E). MEE approaches can be used to explore uncertainty associated with the model time step in **PRA** (e.g., use of a single value to represent a long-term average phenomenon, seasonal patterns in exposure, or intra-individual variability).
Mode	The most probable value of a **random variable**; a value with the largest probability or highest probability density (or mass for discrete random variable). The second **parameter** of a triangular distribution.
Moments of a Distribution	Similar to a parameter; constant that represents a mathematical description of a **random variable**. Central moments are defined with respect to the mean. **Mean**, **Variance**, **Skewness**, and **Kurtosis** are the first, second, third, and fourth central moments of a probability distribution.
Monte Carlo Analysis (MCA) or Simulation	The process of repeatedly sampling from probability distributions to derive a distribution of outcomes (i.e., risks or hazards).
Monte Carlo Sampling	A method of simple random sampling used to obtain a distribution of values which may serve as an input to a probabilistic risk analysis. The probability of obtaining any given sample is similar to the probability of a sample occurring within the distribution. Hence, for a given sample size, simple random sampling tends to produce values clustered around the mean of the distribution.
Multiple Regression Analysis	A statistical method that describes the extent, direction, and strength of the relationship between several (usually continuous) independent variables (e.g., exposure duration, ingestion rate) and a single continuous dependent variable (e.g., risk).
Nonparametric Method	A procedure for making statistical inferences without assuming that the population distribution has any specific form such as normal or lognormal. Sometimes referred to as *distribution-free* methods. Common examples are the sign test, Spearman rank correlation, and the **bootstrap**-t approach.

Definitions of Terms Used in PRA

Numerical stability	The property of a probabilistic simulation such that the a parameter value of the output distribution (e.g., **percentile**, **mean**, **variance**, etc.) remains sufficiently constant for a specified number of **Monte Carlo** iterations (see Chapter 3). Numerical stability is a measure of the precision of the output from a simulation; the tails of the distribution are typically less stable than the center. Sufficient precision is determined by professional judgment.
One-dimensional Monte Carlo Analysis (1-D MCA)	A method of simulating a distribution for an endpoint of concern as a function of probability distributions that characterize variability and/or uncertainty. **PDF**s used in 1-D MCA for human health risks typically represent variability; **PDF**s for ecological risks typically represent uncertainty in the central tendency. It is good practice *not* to combine **PDF**s for variability and uncertainty in 1-D MCA.
Parameter	A constant that characterizes the probability distribution of a **random variable**. For example, a normal probability distribution may be defined by two parameters (e.g., **AM** and **SD**). It is important to distinguish between this definition, and a second popular, but less precise definition that leads to confusion: constants that define a mathematical equation or model. For this guidance, the term ***variable*** will be used to describe the second concept. Body weight is an example of a variable whereas a mean value of 70 kg is an example of a parameter for a distribution of body weight.
Parametric Distribution	A theoretical distribution defined by one or more **parameters.** Examples are the normal distribution, the lognormal distribution, the triangular distribution, and the beta distribution.
Percentile	The p^{th} *percentile* of the distribution is the value such that p percent of the observations fall at or below it. Also called *quantiles* or *fractiles*; percentiles are expressed as a percent, ranging from 0 to 100, quantiles or fractiles range from 0 to 1.
Point Estimate	A quantity calculated from values in a sample to represent an unknown population parameter. Point estimates typically represent a descriptive statistic (e.g., arithmetic **mean**, 95^{th} **percentile**).
Point Estimate Risk Assessment	The familiar risk assessment methodology in which a single estimate of risk is calculated from a set of point estimates. The results provide point estimates of risk for the **CTE** and **RME** exposed individuals. **Variability** and **uncertainty** are discussed in a qualitative manner.
Population-Level effect	An ecological term for an assessment endpoint that focuses on protecting a group of individuals within a specified exposure unit and time that have similar exposures and toxicological responses to chemicals.
Posterior Distribution	A term from **Bayesian** statistics referring to a probability distribution that has been updated with new information.
Power	The probability that a test procedure detects a false null **hypothesis**; Power equals (1 - $), where $ is the probability of a **Type II error** (i.e., accepting H_0 when H_a is true). Power curves are a function of a fixed significance level (**"**), sample size, and **variability** (**SD**).

Definitions of Terms Used in PRA

Preliminary Remediation Goal (PRG)	A health-based chemical concentration in an environmental media associated with a particular exposure scenario. PRGs may be developed based on applicable or relevant and appropriate requirements (**ARARs**), or exposure scenarios evaluated prior to or as a result of the baseline risk assessment.
Prior Distribution	A **Bayesian** term referring to the hypothesized, expected, or calculated probability distribution for an event prior to the collection of new information.
Probabilistic Risk Assessment (PRA)	A risk assessment that uses probabilistic methods to derive a distribution of risk or hazard based on multiple sets of values sampled for random variables.
Probability Density Function (PDF)	A representation, generally a function, graph, or **histogram** (e.g., Fig. 3-1) of the probability of occurrence of an unknown or variable quantity. The sum of the probabilities for discrete random variables, and the integral for continuous random variables, (i.e., the area under the curve) is equal to 1.0. PDFs can be used to display distributions used as input to a probabilistic assessment or the distribution of risks that forms the output of the assessment.
Probability Distribution	A table, graph, or formula that associates probabilities with the values taken by a random variable. Also called a *probability model*.
Random Variable	A variable that may assume any of a set of values. The likelihood of each value is described by a **probability distribution**.
Reasonable Maximum Exposure (RME)	The highest exposure that is reasonably expected to occur at a site (U.S. EPA, 1989; 1990). The intent of the RME is to estimate a conservative exposure case (i.e., well above the average case) that is still within the range of possible exposures.
Reference Dose (RfD)	An estimate of an exposure level for the human population, including sensitive subpopulations, that is likely to be without an appreciable risk of deleterious effects during a lifetime. Chronic RfDs are specifically developed to be protective for a long-term exposure to a chemical (e.g., > 7 years) and accounts for uncertainty spanning perhaps an order of magnitude or greater.
Remediation Action Level	A concentration such that remediation of all concentrations above this level in an exposure unit will result in the **95% UCL** being reduced to a level that does not pose an unacceptable risk to an individual experiencing random exposures. The RAL will depend on the **mean**, **variance**, and sample size of the concentrations within an exposure unit as well as considerations of short term effects of the chemicals of concern.
Remediation Goal	A health-based chemical concentration in an environmental medium chosen by the risk manager as appropriate for a likely land use scenario.
Risk Assessment	The use of available information to make inferences about the health effects associated with exposure of individuals or populations to hazardous materials or situations. Components of risk assessment include: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NAS, 1983).

Definitions of Terms Used in PRA

Risk Characterization	A component of **risk assessment** that describes the nature and magnitude of risk, including uncertainty. In assessments of Superfund sites, it includes the summary and interpretation of information gathered from previous steps in the site risk assessment (e.g., Data Evaluation, Exposure Assessment, Toxicity Assessment), including the results of a probabilistic analysis.
Risk Descriptor	A statistic (e.g., arithmetic **mean**, 95th **percentile**) that describes the risk to the assessment endpoint.
Risk Management	The process by which regulatory decisions are made using all available risk assessment information (including, but not limited to, the results of the PRA). The NCP provides nine criteria for remedial decisions (e.g., protection of human health, conformance with **ARARs**, etc.). Risk managers may include the Remedial Project Manager (RPM), section and branch chiefs, etc.
Scientific/ Management Decision Point (SMDP)	A point during the risk assessment process when the risk assessor communicates results of the assessment at that stage to the risk manager. At this point, the risk manager determines whether the information is sufficient to arrive at a decision regarding **risk management** strategies and/or if additional information is needed to characterize risk.
Sensitivity Analysis	Quantification of the effects of changes in model inputs on model outputs (Chapter 2 and Appendix B). Can be used to rank inputs based on their relative contribution to **variance** in risk. Local sensitivity refers to nominal changes in inputs within a plausible range, whereas range sensitivity refers to changes in inputs across the minimum and maximum values of the plausible range. Definitions for Pearson and Spearman Rank order coefficients are given in Chapter 2.
Sensitivity Ratio	Ratio of the change in model output per unit change in an input variable; also called *elasticity*.
Skewness	The measure of asymmetry of a distribution. Coefficients of skewness are zero for symmetric distributions (e.g., normal), positive for right-skewed distributions (e.g., lognormal), and negative for left-skewed distributions (e.g., specific forms of beta) . Referred to as the third central **moment of a distribution**.
Step Function	A mathematical function that remains constant within each of a series of adjacent intervals but changes in value from one interval to the next. **Cumulative distribution functions** for discrete random variables are step functions.
Stochastic Dominance	Implies no intersection between the **CDFs**; distribution A stochastically dominates distribution B if, for every percentile of the CDF, A > B. This characteristic may not be apparent from the PDFs of the distributions, which may overlap.
Stochastic Process	A process involving random variables, and characterized by variability in space or time.
Time Step	A variable in all exposure models that refers to the unit of time for which a random value is considered representative of intra-individual variability (e.g., average daily ingestion rates for an individual from one year to the next). A time step may be equal to an entire exposure duration (e.g., 30 years), or a fraction of the exposure duration during which changes in input variables may be expected (e.g., one year). Time steps need not be identical for all exposure variables, and should address the most rapidly changing variable in the risk equation. Time step can be an important consideration for **MEE analysis**.

Definitions of Terms Used in PRA

Toxicity Reference Value (TRV)	A numerical expression of a chemical's dose-response relationship that is used in ecological risk assessment.
Truncation	The process of setting lower and upper limits on the range of a distribution, in order to avoid unrealistic values for exposure variables (e.g., > 100% bioavailability). Most often used for continuous, unbounded **probability distributions** (e.g., normal).
Two-dimensional Monte Carlo Analysis (2-D MCA)	An advanced modeling technique that uses two stages of random sampling, also called nested loops, to distinguish between variability and uncertainty in exposure and toxicity variables. The first stage, often called the inner loop, involves a complete **1-D MCA** simulation of variability in risk. In the second stage, often called the outer loop, parameters of the probability distributions are redefined to reflect uncertainty. These loops are repeated many times resulting in multiple risk distributions, from which confidence intervals are calculated to represent uncertainty in the population distribution of risk.
Type I Errors	False positive; the error made when the null hypothesis is rejected in favor of the alternative, when in fact the null hypothesis is true.
Type II Errors	False negative; the error made when the null hypothesis is accepted when in fact the alternative hypothesis is true.
Uncertainty	Lack of knowledge about specific variables, **parameters**, models, or other factors. Examples include limited data regarding the concentration of a contaminant in an environmental medium and lack of information on local fish consumption practices. Uncertainty may be reduced through further study.
Variability	True heterogeneity, diversity, or a range that characterizes an exposure variable or response (e.g., differences in body weight). Further study (e.g., increasing sample size, *n*) will not reduce variability, but it can provide greater confidence in quantitative characterizations of variability.
Variance	Measure of the spread of a distribution, equal to the square of the standard deviation. Variance is referred to as the second central **moment of a distribution**.
Z-score	The value of a normally distributed **random variable** that has been standardized to have a mean of zero and a standard deviation of one by the transformation $Z = (X - \mu)/\sigma$. Statistical tables typically give the area to the left of the z-score value. For example, the area to the left of $z = 1.645$ is 0.95. Z-scores indicate the direction (+/-) and number of standard deviations away from the mean that a particular datum lies assuming X is normally distributed. Microsoft Excel's *NORMSDIST(z)* function gives the probability p such that $p = \Pr(Z \leq z)$, while the *NORMSINV(p)* function gives the z-score z_p associated with probability p such that $p = \Pr(Z \leq z_p)$.

REFERENCES CITED IN APPENDIX A

Ott, W.R. 1995. *Environmental Statistics and Data Analysis.* CRC Lewis, Boca Raton.

U.S. EPA. 1989. *Risk Assessment Guidance for Superfund* (RAGS): *Volume I. Human Health Evaluation Manual* (HHEM) *(Part A, Baseline Risk Assessment).* Interim Final. Office of Emergency and Remedial Response, Washington, DC. EPA/540/1-89/002. NTIS PB90-155581.

U.S. EPA. 1990. National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule. 40 CFR 300: *55 Federal Register*, 8666-8865, March 8.

U.S. EPA. 1991. *Risk Assessment Guidance for Superfund* (RAGS), *Volume 1: Human Health Evaluation Manual* (HHEM), *Part B, Development of Risk-Based Preliminary Remediation Goals.*. Office of Emergency and Remedial Response, Washington, DC. EPA/540/R-92/003. NTIS PB92-963333.

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Office of Solid Waste and Emergency Response, Washington, DC. OSWER Directive 9285.7-081.

U.S. EPA. 1997a. *Memorandum from Deputy Administrator Fred Hansen on the Use of Probabilistic Techniques (including Monte Carlo Analysis) in Risk Assessment, and Guiding Principles for Monte Carlo Analysis.* Office of Research and Development, Washington, DC. EPA/630/R-97/001. May 15, 1997.

U.S. EPA. 1997b. *Lognormal Distribution in Environmental Applications*. Office of Research and Development, and Office of Solid Waste and Emergency Response, Washington, DC. EPA/600/R-97/006. December.

REFERENCES FOR FURTHER READING

Baird, BF. 1989. *Managerial Decisions Under Uncertainty*. John Wiley and Sons, Inc., New York.

Bevington, P.R. 1969. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York.

Bratley, P., B.L. Fox and L.E. Schrage. 1987. *A Guide to Simulation*. Springer-Verlag, New York.

Burmaster, D.E. and P.D. Anderson. 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessment. Risk Anal. 14(4): 477-481.

Clemen, R, 1990. *Making Hard Decisions*. Duxbury Press.

Conover, W.I. 1971. *Practical Nonparametric Statistics*. John Wiley & Sons, New York.

Cox, D.C. and P. Baybutt. 1981. *Methods for Uncertainty Analysis: A Comparative Survey*. Risk Analysis. 1(4): 251-258.

Cullen, A.C. and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment*. Plenum Press, New York.

D'Agostino, R. and M.A. Stephens, eds. 1986. *Goodness-of-Fit Techniques*. Marcel Dekker, Inc., New York.

Devroye, L. 1986. *Non-Uniform Random Deviate Generation*. Springer-Verlag, New York.

Evans, M., N. Hastings and B. Peacock. 1993. *Statistical Distributions*. John Wiley & Sons, New York.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers*. Resources for the Future, Washington, DC.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Hamby, D.M. 1994. A review of Techniques for Parameter Sensitivity Analysis of Environmental Models. Environmental Monitoring and Assessment. 32: 135-154.

Hammersley, J.M. and D.C. Handscomb. 1964. *Monte Carlo Methods*. John Wiley & Sons, New York.

Hertz, D.B. and H. Thomas. 1983. *Risk Analysis and Its Applications*. John Wiley and Sons, New York.

Hertz, D.B. and H. Thomas. 1984. *Practical Risk Analysis - An Approach Through Case Studies*. John Wiley and Sons, New York.

Hoffman, F.O. and J.S. Hammonds. 1992. *An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment*. ES/ER/TM-35, Martin Marietta.

Hoffman, F.O. and J.S. Hammonds. 1994. *Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability*. Risk Analysis. 14(5): 707-712.

Iman, R.L. and W.J. Conover. 1982. *A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables*. Commun. Stat., Part B. 11: 311-331.

Iman, R.L. and J.C. Helton. 1988. *An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models.* Risk Analysis. 8(1): 71-90.

Iman, R.L., J.M. Davenport and D.K. Zeigler. 1980. *Latin Hypercube Sampling (A Program Users Guide)*. Technical Report SAND 79:1473, Sandia Laboratories, Albuquerque.

Johnson, M.E. 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, New York.

Johnson, N.L. and S. Kotz. 1970. *Continuous Univariate Distributions*. Vols. 1 & 2. John Wiley & Sons, New York.

Johnson, N.L., S. Kotz and A.W. Kemp. 1992. *Univariate Discrete Distributions*. John Wiley & Sons, New York.

Kendall, M. and A. Stuart. 1979. *Advanced Theory of Statistics, Volume I - Distribution Theory, Volume II - Inference and Relationship*. MacMillan, Inc., New York.

Kennedy, W.J. and E. Gentle. 1980. *Statistical Computing*. Marcel Dekker, Inc., New York.

Law, A.M. and W.D. Kelton. 1991. *Simulation Modeling and Analysis.* McGraw-Hill, Inc., New York.

LePage, R. and L. Billard. 1992. *Exploring the Limits of Bootstrap*. John Wiley & Sons, New York.

Lipton, J., Shaw, W.D., Holmes, J., and A. Patterson. 1995. Short communication: selecting input distributions for use in Monte Carlo analysis. *Regulatory Tox. and Pharm*. 21: 192-198.

McKone, T.E. and K.T. Bogen. 1992. Uncertainties in health risk assessment: an integrated case based on tetrachloroethylene in California groundwater. *Regulatory Tox. and Pharm*. 15: 86-103.

Megill, R.E., ed. 1985. *Evaluating and Managing Risk*. Penn Well Books, Tulsa, Oklahoma.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, New York.

NCRP. 1996. Commentary No. 14. *A Guide for Uncertainty Analysis in dose and Risk Assessments Related to Environmental Contamination*. National Committee on Radiation Programs, Scientific Committee 64-17. Washington, DC.

Palisade Corporation. 1994. *Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3*. Windows Version Release 3.0 User's Guide, Palisade Corporation, Newfield, New York.

Press, W.H., B.P. Flannery, S.A. Teulolsky and W.T. Vetterling. 1989. *Numerical Recipes in Pascal: the Art of Scientific Computing*. Cambridge University Press, New York.

Press, W.H., S.A. Teulolsky, W.T. Vetterling and B.P. Flannery. 1992. *Numerical Recipes in FORTRAN: the Art of Scientific Computing*. Cambridge University Press, New York.

Press, W.H., S.A. Teulolsky, W.T. Vetterling and B.P. Flannery. 1992. *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, New York.

Read, T. and N. Cressie. 1988. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, New York.

Rohatgi, V.K. 1984. *Statistical Inference*. John Wiley & Sons, New York.

Rubenstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. John Wiley and Sons, New York.

Sachs, L. 1984. *Applied Statistics - A Handbook of Techniques*. Springer-Verlag, New York.

Saltelli, A and J. Marivort. 1990. *Non-Parametric Statistics in Sensitivity analysis for Model Output: A Comparison of Selected Techniques*. Reliability Engineering and System Safety. 28: 299-253.

Schneider, H. 1986. *Truncated and Censored Distributions from Normal Populations*. Marcel Dekker, Inc., New York.

Seiler, F.A. 1987. Error propagation for large errors. *Risk Anal*. 7(4): 509-518.

Seiler, F.A. and J.L. Alvarez. 1996. On the selection of distributions for stochastic variables. *Risk Anal*. 16(1): 5-18.

Slob, W. 1994. Uncertainty analysis in multiplicative models. *Risk Anal*. 14(4): 571-576.

Smith, A.E., P.B. Ryan and J.S. Evans. 1992. The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Anal*. 12(4): 467-474.

Smith, R.L.. 1994. Uses of Monte Carlo simulation for human exposure assessment at a Superfund site. *Risk Anal*. 14(4): 433-439.

Sokal, R. and R. Rohlf. 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*. Second Edition. W.H. Freeman & Co., New York.

U.S. EPA. 1978. *Source Assessment: Analysis of Uncertainty - Principles and Applications*. EPA/600/2-79-004.

U.S. EPA. 1992a. *Guidelines for Exposure Assessment*. Federal Register. May 29, 1992. 57(104): 22888-22938.

U.S. EPA. 1992b. *Guidelines for Carcinogenic Risk Assessment*. Federal Register. May 29, 1992. 51(185): 33992-34003.

U.S. EPA. 1996. *Summary Report for the Workshop on Monte Carlo Analysis.* Office of Research and Development, Washington, DC. EPA/630/R-96/010.

Wilks, D.S. 1995. *Statistical Methods in the Atmospheric Sciences, An Introduction*. Academic Press, San Diego.

APPENDIX B

ADVANCED CONCEPTS IN SENSITIVITY ANALYSIS

As described in Chapter 2, there are several approaches to sensitivity analysis that may be useful in probabilistic risk assessment. The basic concept of a sensitivity analysis is to understand how risk estimates are influenced by variability and uncertainty in the risk model. While statistical software for Monte Carlo analysis provides convenient metrics for quantifying and ranking these sources, it is strongly recommended that risk assessors and risk managers develop an understanding of the underlying principles associated with these metrics. This appendix provides additional information on the underlying principles of sensitivity analysis, although it is not a comprehensive summary and is not intended to substitute for the numerous statistical texts and journal articles on sensitivity analysis. Section B.1 begins with a general framework for relating model output to model input. Section B.2 explains the sensitivity ratio approach and highlights some of its limitations. Section B.3 reviews some of the metrics reported by the commercial software that report results of sensitivity analysis following Monte Carlo simulations (e.g., Crystal Ball®, @Risk®).

B.1 RELATING THE CHANGE IN RISK TO THE CHANGE IN EXPOSURE VARIABLE X

For purposes of discussion, let Y denote a model output (e.g., risk) and suppose that it depends on two input variables denoted by X_1 and X_2. In general, a risk assessment model may use any number of inputs; however, for purposes of illustrating concepts, it is convenient to restrict this discussion to two variables. The model relates the output Y to the two inputs through a function expressed as $Y = f(X_1, X_2)$. This function represents a surface in three dimensional space where the vertical axis represents Y and the two horizontal axes represent inputs X_1 and X_2 (Figure B-1a). The form or "shape " of this surface may be very simple, such as a plane (Figure B-1b), or it may be very complex with many "hills" and "valleys" depending on the defining function $f(X_1, X_2)$ (Figure B-1a). The way in which Y changes in response to a change in either X_1 or X_2 may, therefore, depend on where we are on this surface. In local sensitivity analysis, the objective is to evaluate the sensitivity at some nominal point (X_1^*, X_2^*) such as the point defined by the mean or median of X_1 and X_2. At that point, or at any other point for that matter, the sensitivity of the model output Y (Y^*) to one of the inputs, X_i, is represented by the rate of change in Y per unit change in X_i. This is the slope of the surface at that nominal point in the direction of X_i and is expressed as $\partial Y/\partial X_i$, the **partial derivative** of Y with respect to X_i.

$$Partial\ Derivative = \frac{\partial Y}{\partial X_i} \approx \frac{\Delta Y}{\Delta X_i}$$

If the function $f(X_1, X_2)$ is known explicitly, it may be possible to determine the partial derivatives analytically. This is not a requirement, however, because an estimate can be obtained by incrementing X_i by a small amount, ΔX_i, while keeping the other inputs fixed and reevaluating the model output Y. The

resulting change in Y divided by ΔX_i will approximate $\partial Y/\partial X_i$ at the nominal point. In practice, analytical solutions can be approximated using Monte Carlo techniques.

One drawback to using the partial derivative to quantify the influence of X_i is that the partial derivative is influenced by the units of measurement of X_i. For example, if the measurement scale for X_i is changed from grams to milligrams, the partial derivative $\partial Y/\partial X_i$ will change by a factor of 1000. Therefore, it is necessary to **normalize the partial derivative** to remove the effects of units (see Appendix B.3).

If the relationship between Y and all of the inputs is linear, then the response surface is a flat plane and each of the partial derivatives at each point, (X_i, Y), will remain constant regardless of where the point is in the surface (Figure B-1b). In this case, it is a simple matter to determine the relative influence that the various inputs have on the model output. When the relationship is nonlinear, however, the situation is more complex because the influence of a particular input may vary depending on the value of that input.

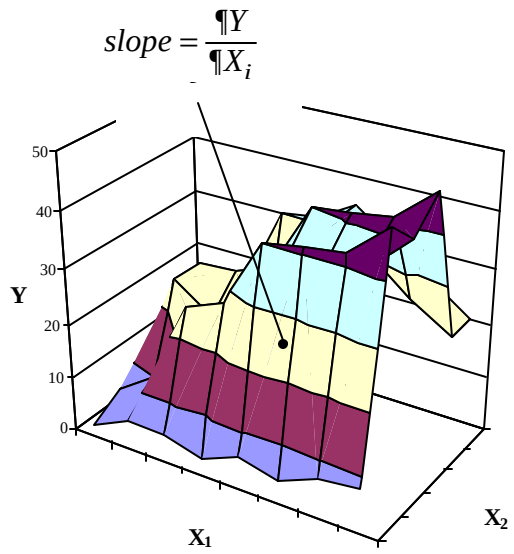


Figure B-1a. Hypothetical 3-D response surface for *Y* given two input variables: Y = *f*(X_1, X_2). The sensitivity of *Y* with respect to X_i is calculated as the slope at a specific point on the surface, or the partial derivative, $\partial Y/\partial X_i$.

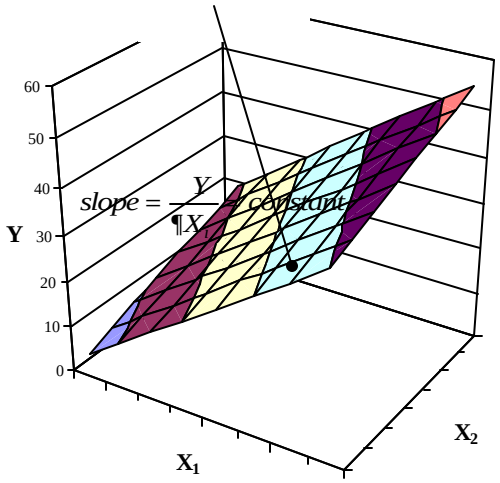


Figure B-1b. Hypothetical 3-D response surface when *Y* is a linear function of two input variables: Y = *f*(X_1, X_2). The slope (i.e., the partial derivative, $\partial Y/\partial X_i$) is constant for any point (X_i, *Y*) on the surface in the direction of X_i. In this case, $\partial Y/\partial X_1$ = 5 while $\partial Y/\partial X_2$ = 2.

B.2 CALCULATING SENSITIVITY RATIOS (BOTH LOCAL AND RANGE)

Sensitivity ratios can be used for sensitivity analysis in both point estimate and probabilistic risk assessment. The approach is easy to understand and apply. The ratio is equal to the percentage change in output (e.g., risk) divided by the percentage change in input for a specific input variable (see Appendix Equation. B-1). Risk estimates are considered most sensitive to input variables that yield the highest ratios. For simple exposure models in which the relationship between exposure and risk is linear, the ratio offers little information regarding the relative contributions of each input variable to the variance in risk. However, for more complex, nonlinear models, the ratio can offer a useful screening tool to identify potentially influential input variables.

Sensitivity ratios can generally be grouped into two categories. For the local sensitivity ratio method, an input variable is varied by a small amount, usually ±5% of the nominal (default) point estimate, and the corresponding change in the model output is observed. For the range sensitivity ratio method, an input variable is varied across the entire range (plausible minimum and maximum values). If local and range sensitivity approaches yield different results, the risk assessor can conclude that different exposure variables are driving risk near the high-end (i.e., extreme tails of the risk distribution) than at the central tendency region. Equation B-3 can be used to evaluate SR for different types of exposure models in which the intake equation is generally expressed as a simple algebraic combination of input variables.

One difficulty with sensitivity ratio approach is that the method assumes that the input variables are independent. If two exposure variables are correlated, holding the value for one fixed while allowing the value for the other to vary may produce misleading results, especially with the range sensitivity ratio approach. For example, it may not be realistic to hold body weight fixed at a central tendency while allowing skin surface area to vary from the minimum to maximum values. An improvement over the sensitivity ratio approach would be to allow correlated input variables to vary simultaneously.

The general equation for calculating a Sensitivity Ratio (SR) for a variable is given by Equation. B-1.

$$SR = \frac{\left(\frac{D_2 - D_1}{D_1}\right) \times 100\%}{\left(\frac{V_2 - V_1}{V_1}\right) \times 100\%} \qquad \text{Equation B-1}$$

where,

D_1 = the value of the output variable using unchanged values of input variables

D_2 = the value of the output variable after changing the value of one input variable

V_1 = the default point estimate for an input variable

V_2 = the value of the input variable after changing V_1

Let) equal the percentage change in the input variable, V_1. For *SR* calculations,) may be either positive or negative (e.g., ±5% for local SR; ±100 for range SR), and the new value for the input variable (i.e., V_2) is given by Equation. B-2.

$$V_2 = V_1 + (V_1 \cdot \Delta) = V_1 \cdot (1 + \Delta) \qquad \text{Equation B-2}$$

Therefore, the denominator in Equation B-1 reduces to):

$$\frac{V_2 - V_1}{V_1} = \frac{V_1(1+\Delta) - V_1}{V_1} = \frac{(1+\Delta) - 1}{1} = \Delta$$

and Equation B-1 reduces to Equation B-3:

$$SR = \frac{1}{\Delta} \cdot \left(\frac{D_2 - D_1}{D_1} \right) \qquad \text{Equation B-3}$$

Equation B-3 can be used to evaluate SR for different types of exposure models in which the intake equation is generally expressed as a simple algebraic combination of input variables (for examples, refer to Chapter 2, and RAGS Vol. 3 Part B).

B.2.1 NORMALIZING THE PARTIAL DERIVATIVE

Classical sensitivity analysis methods use estimates of the partial derivatives of the model output with respect to each variable (Tomovic, 1963). For the purpose of evaluating the relative influence of the various input variables on the model output at a single point, the **normalized partial derivative** provides a useful index.

If the input variables are all discrete and take on a small number of values, then it is possible to evaluate the influence of the various input variables at each of the points defined by considering all possible combinations of the inputs. Then the influence can be evaluated for each input by computing normalized partial derivatives at each point. This approach is limited to situations where the number of inputs as well as the number of possible values for each input is relatively small; otherwise, the number of combinations to be evaluated will be unmanageable. Furthermore, when evaluating the influence at

different points on the input-output surface simultaneously, it is important to take into account the probability associated with each of those points. For example, the fact that a particular input has a large influence on the model output at a particular point would be discounted if the probability associated with that particular point is very low.

A similar approach may be used to analyze inputs that are continuous variables if a few points representing the range of values are selected. For example, low, medium (or nominal), and high values may be selected for each of the continuous input variables and then the relative influence of each of the input variables can be computed as in the case of discrete inputs. One limitation of this approach, however, is that the continuous nature of the inputs makes it impossible to calculate an exact probability for each of the points. Generally, in a PRA, many if not all of the inputs will be random variables described by probability distributions and it will be necessary to quantify the influence of each input, X_i, over the entire range of X_i.

An estimate of the partial derivative can be obtained by incrementing X_i by a small amount, say ΔX_i while keeping the other inputs fixed and reevaluating the model output Y. The resulting change in Y divided by ΔX_i will approximate $\partial Y/\partial X_i$ at the nominal point.

$$\text{Partial Derivative} = \frac{\partial Y}{\partial X_i} \approx \frac{\Delta Y}{\Delta X_i}$$

As previously noted, one complication to using the partial derivative to quantify the influence of X_i is that the partial derivative is influenced by the units of measurement of X_i. One way this is accomplished is to divide the partial derivative by Y^* / X_i^* (or equivalently multiply by X_i^* / Y^*). An approximation of the normalized partial derivative is given below.

$$\text{Normalized Partial Derivative} = \frac{\left(\frac{\Delta Y}{\Delta X_i}\right)}{\left(\frac{Y_1}{X_{i,1}}\right)} \approx \frac{\left(\frac{Y_1 - Y_2}{Y_1}\right)}{\left(\frac{X_1 - X_2}{X_1}\right)_i}$$

Another approach is to divide by the ratio (σ_Y / σ_X), where σ_Y is the standard deviation of Y and σ_X is the standard deviation of X. The latter method requires that the standard deviations be known, or that a suitable estimate can be obtained.

As previously noted, if the relationship between *Y* and all of the inputs is nonlinear, the influence of a particular input may vary depending on the value of that input. One approach to this problem is to consider a range of values for the input and to examine the influence over that range. If the input is considered to be a random variable following some specified probability distribution, then it may be desirable to look at the influence that the random input has on the model output across the distribution of input values. This can be accomplished with a Monte Carlo approach.

B.3 REGRESSION ANALYSIS: R^2, PEARSON R, AND PARTIAL CORRELATION COEFFICIENTS

In order to understand R^2, it is necessary to first understand simple and multiple linear regression. In regression analysis, we are interested in obtaining an equation that relates a dependent variable (*Y*) to one or more independent variables (*X*):

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{Equation B-4}$$

where β_0 and β_1 are regression coefficients, and ε is called a random error. Equation B-4 is the general equation for a simple linear regression, because we have only one Y and one X variable, and their relationship can be described by a line with intercept β_0 and slope β_1. Note that *linear* regression refers to the linear relationship between parameters (β_0, β_1), not *X* and *Y*; thus the equation $Y = \beta_0 + \beta_1 X_1^2 + \varepsilon$ is considered linear. *Multiple* linear regression involves more than one *X* related to one *Y* $[Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots]$, while *multivariate* regression involves more than one *Y* to more than one *X*.

The random error, ε, represents the difference between an observed *Y* value (calculated from the observed input variables), and a Y value predicted by the regression line ($\hat{y}$). It is also called the *residual* (i.e., $\varepsilon = y - \hat{y}$). The random error takes into account all unpredictable and unknown factors that are not included in the model. Assumptions about ε are that the random error has mean = 0 and constant variance, and is uncorrelated among observations. One method of finding the best regression line is to minimize the residual sum of squares (i.e., least-squares method), also called the sum of squares due to error (SSE).

In terms of sensitivity analysis, we are interested in how much of the variation in *Y* can be explained by the variation in *X*, and how much is unexplained (due to random error). If a scatterplot of paired observations (*x*, *y*) shows that our regression line intersects all of the observations exactly, then all of the variation in *Y* is explained by *X*. Another way of stating this is that the difference between the mean output ($\bar{y}$) and an observed *y* (y_i), or ($y_i - \bar{y}$), is equal to the difference between the mean output ($\bar{y}$) and a predicted *y* ($\hat{y}_i$), or ($\hat{y}_i - \bar{y}$).

Simplifying Assumptions in Regression Analysis

- *Y* is a linear function of the unknown coefficients (β_i)
- successive values of *Y* are uncorrelated
- variance of *Y* is constant for all values of inputs (X_i)

In general, the total deviation of y_i from $\bar{y}$ is equal to the sum of the deviation due to the regression line plus the deviation due to random error:

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \qquad \text{Equation B-5}$$

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$$

$$SST = SSE + SSR$$

Thus, the total sum of squares (SST) equals the sum of squares due to error (SSE) plus the sum of squares due to regression (SSR).

B.3.1 CALCULATIONS OF R^2 AND ADJUSTED R^2

The R^2 term is a measure of how well the regression line explains the variation in *Y*, or:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad \text{Equation B-6}$$

$$R = \sqrt{\frac{\text{variation explained by regression}}{\text{Total variation in } Y}}$$

where R^2 is called the *coefficient of multiple determination* and R is called the *multiple correlation coefficient*. If $R^2 = 0.90$ for a certain linear model, we could conclude that the input variables (X_1, X_2,...X_k) explain 90% of the variation in the output variable (*Y*). R^2 reduces to the *coefficient of determination* r^2 for simple linear regression when one independent variable (*X*) is in the regression model. The *sample correlation coefficient, r,* is a measure of the association between *X* and Y, and calculated by Equation. B-7.

Equation B-7

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2\right]^{0.5}}$$

In addition, *r* is an estimate of the unknown population parameter, ρ, defined by Equation. B-8:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad \text{Equation B-8}$$

where σ_X and σ_Y denote the population standard deviations of the random variables *X* and *Y*, and where σ_{XY} is called the covariance between *X* and *Y*. The covariance σ_{XY} is a population parameter describing the average amount that two variables "covary". Thus, another way of thinking about a correlation coefficient (*R*) is that it reflects the ratio of the covariance between two variables divided by the product of their respective standard deviations; and the value always lies between -1 and +1. *@Risk* and *Crystal Ball* provide both the R^2 for the entire model, as well as the correlation coefficients for each input variable (or regressor). The higher the value of R_i for X_i, the more sensitive the output variable is to that input variable.

Although the calculations are the same, there is a subtle conceptual difference between the coefficient of determination (r^2) from regression, and the square of the correlation coefficient. When evaluating two variables [X, Y], the key is whether X is interpreted as a "fixed" quantity (i.e., an explanatory variable), or a random variable just like Y. In regression analysis, r^2 measures how well the regression line explains the variation in Y given a particular value for X (Equation B-6). Correlation requires that X be considered a random variable, typically having a bivariate normal distribution with Y (see Chapter 3).

One artifact of regression analysis is that R^2 increases as you add more and more input variables to your model; however, the increased fit of the model due to one or more of the input variables may be insignificant. Sometimes an adjusted R^2 is calculated to take into account the number of input variables in the model (k) as well as the number of observations in the data set (n):

$$R^2_{adj} = \frac{(n-1)R^2 - 1}{n - k - 1} \qquad \text{Equation B-9}$$

While R^2 gives the proportion of the total *variation* of *Y* that is explained, R^2_{adj} takes into account the degrees of freedom (*df*), and gives the proportion of the total *variance* of Y that is explained (variance = variation /df); or stated simply, R^2_{adj} is the R^2 corrected for *df*.

- If the relationship between an input variable and an output variable is strong, but nonlinear, the R^2 statistic will be misleadingly low.

- If the means of the sampling data are used rather than the individual observations for each variable, R^2 will be misleadingly high. This is because taking the mean of a sample reduces the fraction of the *total* variation due to *random* variation (see discussion of random error above). This is an important consideration when trying to interpret the results of regression analyses that incorporate data averaged over different spatial scales (e.g., regression of PbB on soil lead concentrations taken at the city block level may give an inflated R^2 value if the sampling data are averaged over a larger spatial scale, such as the census tract level).

A multiple regression analysis can also be performed to estimate the **regression coefficients** (see Appendix B.3). Each coefficient essentially represents an "average" value of the partial derivative across the entire distribution of the input. The regression coefficient, like the partial derivative, depends on the units of measurement so, as in the case of the partial derivative, it must be normalized. This can be accomplished by multiplying the regression coefficient by the ratio of estimated standard deviations s_y/s_x.

A convenient way to carry out a sensitivity analysis is to perform a stepwise regression analysis. Some statistical software packages (e.g., SAS, SPSS) offer a variety of different approaches for this, however, in general, they can be classified into two general categories: forward selection and backward elimination. In the forward selection, the inputs are added to the model one by one in the order of their

contribution. In the backward elimination, all of the inputs are used in the model initially and then they are dropped one by one, eliminating the least important input at each step. A true stepwise procedure is a variation on the forward selection approach where an input can drop out again once it has been selected into the model if at some point other inputs enter the model that account for the same information.

B.3.2 RELATIVE PARTIAL SUM OF SQUARES (RPSS)

The **relative partial sum of squares (RPSS)** measures the sensitivity of the model output to each of the input variables by partitioning the variance in the output attributable to each variable using multiple regression techniques (Rose et al., 1991). The RPSS is presented as a percentage reflecting the proportion of influence a given variable has on risk. The results of RPSS are intuitive and generally easy to understand.

Briefly, the RPSS represents the percentage of the total sum of squares attributable to each of the variables. To calculate RPSS for variable V_i, the difference between the regression sum of squares (RSS) for the full model and the regression sum of squares for the model with V_i missing (RSS_{-i}) is divided by the total sum of squares (TSS) and expressed as a percentage:

$$RPSS_i = \frac{100 \cdot (RSS - RSS_{-i})}{TSS} \qquad \text{Equation B-10}$$

This procedure can be thought of as analogous to least squares linear regression but performed in the *n*-dimensional parameter space of the risk equation. Since this approach depends on the adequacy of the linear regression model between the output variable (e.g., risk) and all the variables, an additional diagnostic is to check how close R^2 is to 1.0. For equations with more than three parameters (such as those used in Superfund risk assessments), the computational overhead of this process is large and requires specific computer programs. The software program *Crystal Ball* does not perform this calculation, but it can be determined with most standard statistical software packages that perform multiple regression. *@Risk* performs a calculation similar to this called multivariate stepwise regression that yields correlation coefficients in lieu of percent contributions to output variance.

B.3.3 SPEARMAN'S RANK CORRELATION COEFFICIENT (RHO)

The validity of using indices such as regression coefficients, correlation coefficients, and partial correlation coefficients depends on the assumptions of the underlying linear model being met. If there is any doubt that a data set satisfies the model assumptions, a nonparametric measure of correlation based on the rank orders of the inputs and associated outputs can be used. The Spearman Rank correlation coefficient is a non-parametric statistic; it measures an association between variables that are either count data or data measured on an ordinal scale, as opposed to data measured on an interval or ratio scale. An example of an ordinal scale would be the ranking of sites based on their relative mean soil concentrations -- for example, if there are four categories of soil contaminant concentrations, sites with the highest

concentrations may receive a rank of 1 while sites with lowest concentrations may receive a rank of 4. Ordinal scales indicate relative positions in an ordered series, not "how much" of a difference exists between successive positions on a scale.

To calculate the Spearman rank correlation coefficient, assign a rank to each of the input variables (X_j) and output variables (Y_k). For each ranked pair (X_j, Y_k), calculate the difference, d, between the ranks. For example, if the first observation for variable X has a ranking of 5 (relative to all of the observations of X), and the corresponding value of Y has a ranking of 3 (relative to all of the observations of Y), the difference (d) is equal to 5 - 3 = 2. Spearman rho (r_s) is calculated as:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{(n^3 - n)}$$

Equation B-11

Hence -1 # r_s # 1.0, and r_s = -1 describes a perfect indirect or negative relationship between ranks in the sense that if an X element increases, the corresponding Y element decreases. Similarly, r_s = 0 suggests that there is no relationship between X and Y.

The Pearson product moment correlation coefficient is equal to the Spearman rank correlation coefficient when interval/ratio values of the measured observations (X, Y) are replaced with their respective ranks.

Some statistical packages offer the correlation coefficient as an index of sensitivity, so it is important to identify if a parametric or nonparametric measure is being used. *Crystal Ball* and *@Risk* can be used to calculate the Spearman Rank Correlation. Rank correlation coefficients shown in *Crystal Ball* and *@Risk* are calculated by the standard method provided in most statistics texts. *Crystal Ball* also indicates that sensitivity can be determined as contribution to variance. This is not the relative partial sum of squares techniques discussed above. Instead, *Crystal Ball* calculates the contribution to the variance by squaring the rank correlation coefficients and normalizing them to 100%. Many other commonly used commercial software packages will perform Spearman Rank Correlation.

APPENDIX C

PROBABILITY DISTRIBUTIONS FOR PRA

C.0 INTRODUCTION

Once a determination has been made that variability or uncertainty will be characterized by a probability density function (PDF), a risk assessor has a wide range of options to select from. Chapter 3 provides guidance on factors to consider in selecting and fitting PDFs to data. The choice of distribution family and parameter estimates will depend on many factors. Often, more than one PDF may adequately characterize variability, and judgment is required to define plausible bounds. Uncertainty in exposure variables may be described based on a statistical measure of parameter uncertainty (e.g., upper and lower 95% confidence limits), or an evaluation of parameter estimates from data sets describing surrogate populations.

Table C-1 lists some of the probability distributions that may commonly be used in PRA. This is not an exhaustive list, and the scientific literature contains numerous examples with alternative distributions. Where practicable, a mechanistic basis is presented for the choice of the distribution. For some distributions, such as beta, triangular, and uniform, a mechanistic basis is not offered because it is unlikely that a chemical or biological process will yield a random variable with that particular distribution. Nevertheless, such distributions may be appropriate for use in PRA because they reflect the extent of information that is available to characterize a specific random variable. Because of there flexible shapes, they may be useful models of variability for a variety of exposure variables.

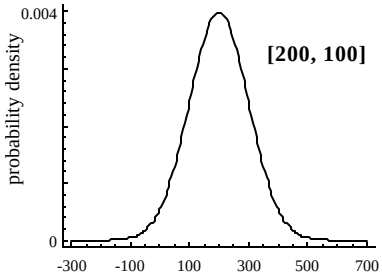
Figure C-1 presents examples of PDFs and the corresponding CDFs for the following distributions: a) normal; b) lognormal; c) triangular; d) uniform; e) beta; f) Weibull; g) gamma; and h) exponential. For each distribution, one or more examples with different parameter estimates are given to demonstrate the flexibility in the shape of the PDF. See Table 3-1 for a summary of the parameters and theoretical bounds that define the PDFs.

Table C-1 - Examples of Selected Probability Distributions for PRA

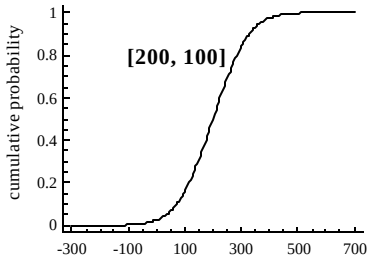
Distribution	Mechanistic Basis	Example(s)
Beta	Describes a continuous random variable with finite upper and lower bounds. This distribution can take on very flexible shapes, but generally does not have a mechanistic basis.	Absorption fraction bounded by 0 and 100%; fraction of time an individual spends indoors.
Binomial	Describes a discrete random variable produced by processes that (1) occur in a fixed number *n* of repeated independent "trials", (2) yield only one of two possible outcomes (i.e., "success" or "failure") at each trial, and (3) have constant probability *p* of "success". A beta distribution is characterized by parameters *n*, *p*, and *x*, representing the number of trials, the probability of success of each trial, and the number of successes, respectively.	The number of animals with tumors (or some other quantitative outcome) in a chronic animal bioassay.
Exponential	Results if instead of counting the number of events in the Poisson process (below), one measures the time (or distance) between any two successive events.	The length of time between two radiation counts; length of time between major storm events; distance between impact points of two artillery shells.
Gamma	Similar to exponential except that time until occurrence of the k^{th} event in the Poisson process is measured (rather than time between successive events). Reduces to exponential when k = 1.	Time until k^{th} radiation count; elapsed time until k^{th} major storm event.
Lognormal	Multiplication of a large number of random variables, or equivalently adding the logarithms of those numbers, will tend to yield a distribution with a lognormal shape (i.e., the logarithms of the products which is the sum of the logarithms will be normally distributed by the central limit theorem).	Chemical concentrations in environmental media; media contact rates; rates and flows in both fate and transport models. Because the basic risk equation is multiplicative, distributions of risk are generally lognormal. In practice, lognormal distributions often provide good fits to data on toxicant concentrations in a variety of media (Gilbert, 1987; Ott, 1990).

Distribution	Mechanistic Basis	Example(s)
Normal	Addition of independent random variables, with no one variable contributing substantially to the total variation of the sum, will tend to yield a distribution with a normal shape. This result is established by a basic theorem of probability theory (i.e., central limit theorem).	The "Gaussian Plume Model" for the dispersion of air pollutants is based on the idea that, at a micro level, individual parcels of air, or molecules of pollutants, are subject to many random collisions from other molecules that act together as if a large number of random numbers were being added/subtracted from an initial 3-dimensional description of a position.
Poisson	Observed when counting the frequency of discrete events, where the events are independent of one another, and randomly distributed in space or time. Approximates the binomial distribution when n is large and p is small.	The number of counts of radiation that occur in a particular time interval; the release of synaptic transmitter from nerve cells; the number of artillery shells falling within a fixed radius; the occurrence of major storm events in a month; number of leaks in average length of pipe.
Triangular	The PDF is shaped like a triangle, with parameters representing plausible bounds and a most likely value (i.e., mode). This is a "rough" probability model that generally describes a random variable based on limited information rather than mechanistic basis.	Variability in shower droplet diameter. Uncertainty in the mean air exchange rate in a shower.
Uniform	The PDF is shaped like a rectangle, with parameters representing plausible bounds. This is a "rough" probability model that generally describes a random variable based on limited information rather than a mechanistic basis.	Variability in the air ventilation rate in a house.
Weibull	Originated in reliability and (product) life testing as a model for time to failure or life length of a component when the failure rate changes with time. A very flexible model taking a wide range of shapes. If the failure rate is constant with time, the Weibull reduces to the exponential distribution.	Examples for exponential and gamma would also be appropriate for Weibull.

Normal

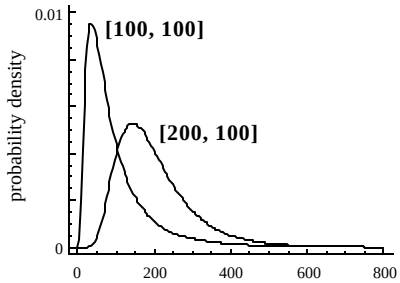


(a1)

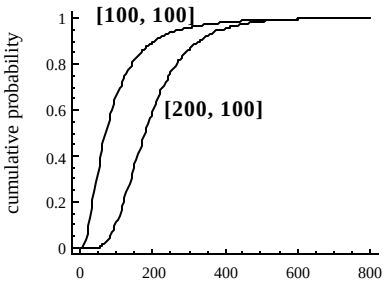


(a2)

Lognormal

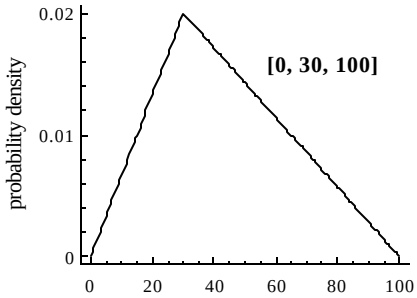


(b1)

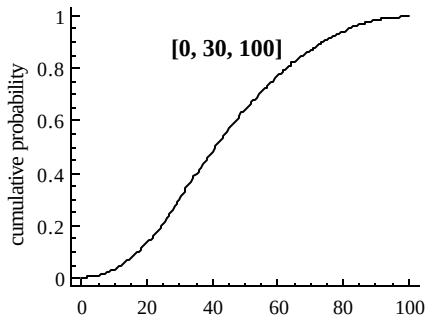


(b2)

Triangular

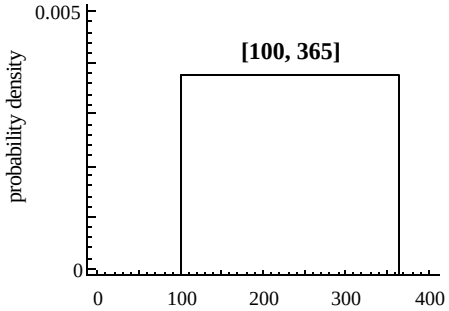


(c1)

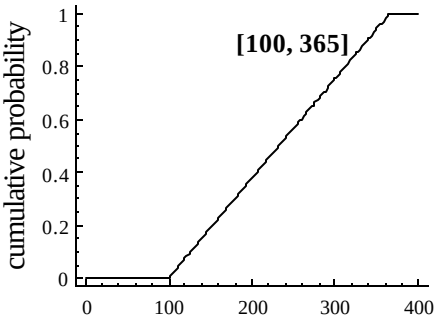


(c2)

Uniform

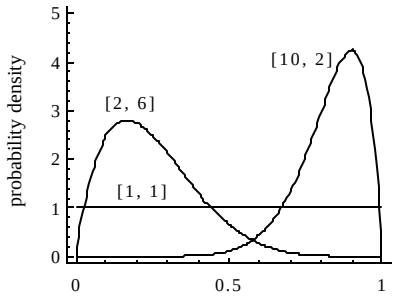


(d1)

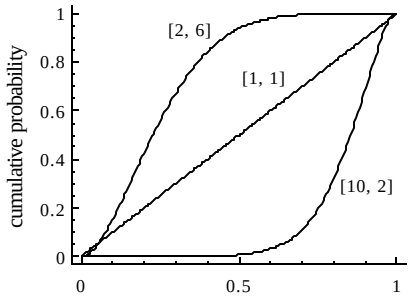


(d2)

Beta

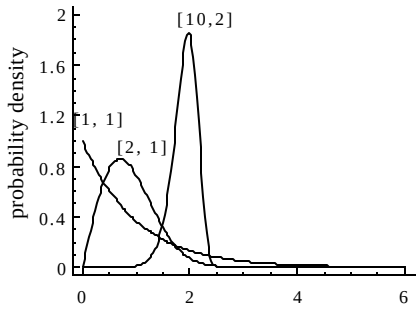


(e1)

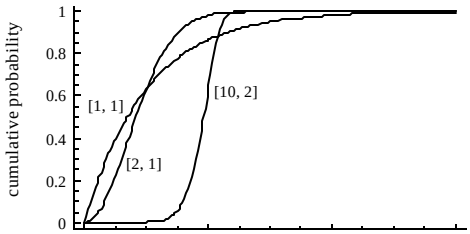


(e2)

Weibull

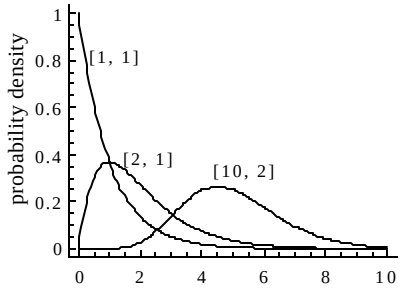


(f1)

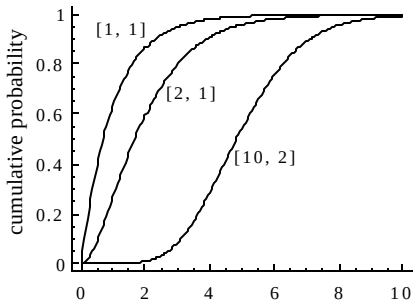


(f2)

Gamma

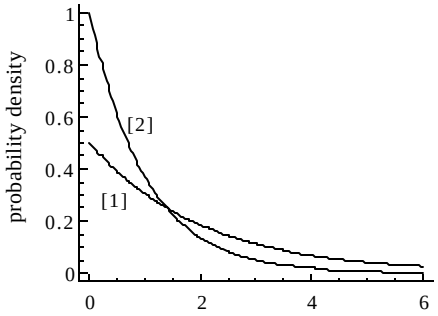


(g1)

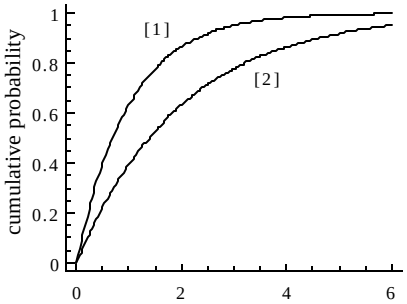


(g2)

Exponential



(h1)



(h2)

APPENDIX D (PART 1 OF 2)

ESTIMATING UNCERTAINTY IN THE MEAN CONCENTRATION

D.0 INTRODUCTION

An important step in characterizing site-specific risks is the derivation of the concentration term, which should be used in the intake equation for Superfund exposure assessments (U.S. EPA, 1989). The general equation used for calculating exposure is shown in Equation D-1.

In a point estimate risk assessment, the concentration term is an estimate of the arithmetic mean concentration within the exposure unit for a contaminant based on a sample of observations made at the site. Because of the uncertainty associated with estimating the true mean concentration at a site when data are limited, EPA recommends using the 95 percent upper confidence limit (95% UCL) of the arithmetic mean for this variable. EPA's guidance on calculating the concentration term describes the rationale and methodology for selecting this point estimate as an input variable (U.S. EPA, 1992). Chapter 4 discusses how to develop a PDF for the concentration term for use in PRA in a manner that is consistent with the objective of estimating long-term risks. This appendix provides additional details regarding two common approaches for characterizing uncertainty in the arithmetic mean concentration: (1) the Land method for lognormal distributions; and (2) the bootstrap resampling techniques.

EQUATION D-1

GENERAL EQUATION FOR ESTIMATING EXPOSURE TO A SITE CONTAMINANT

$$I = \frac{C \cdot CR \cdot EF \cdot ED}{BW \cdot AT}$$

where,

I	=	daily intake
C	=	contaminant concentration
CR	=	contact rate (ingestion, inhalation, dermal contact)
EF	=	exposure frequency
ED	=	exposure duration
BW	=	body weight
AT	=	averaging time

D.1 LAND METHOD FOR CALCULATING THE 95% UCL FOR THE ARITHMETIC MEAN OF A LOGNORMAL PDF

The Land method refers to the conventional method for calculating the 95% UCL for the mean concentration when the sample data are obtained from a lognormal distribution of concentrations. EPA's *Supplemental Guidance for RAGS: Calculating the Concentration Term* (U.S. EPA, 1992) provides details and examples of applications of the Land method. The method is

EQUATION D-2

CALCULATING THE 95% UCL FOR THE AM OF A LOGNORMAL DISTRIBUTION

$$95\%\,UCL = \exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{0.95}}{\sqrt{n-1}}\right)$$

where,

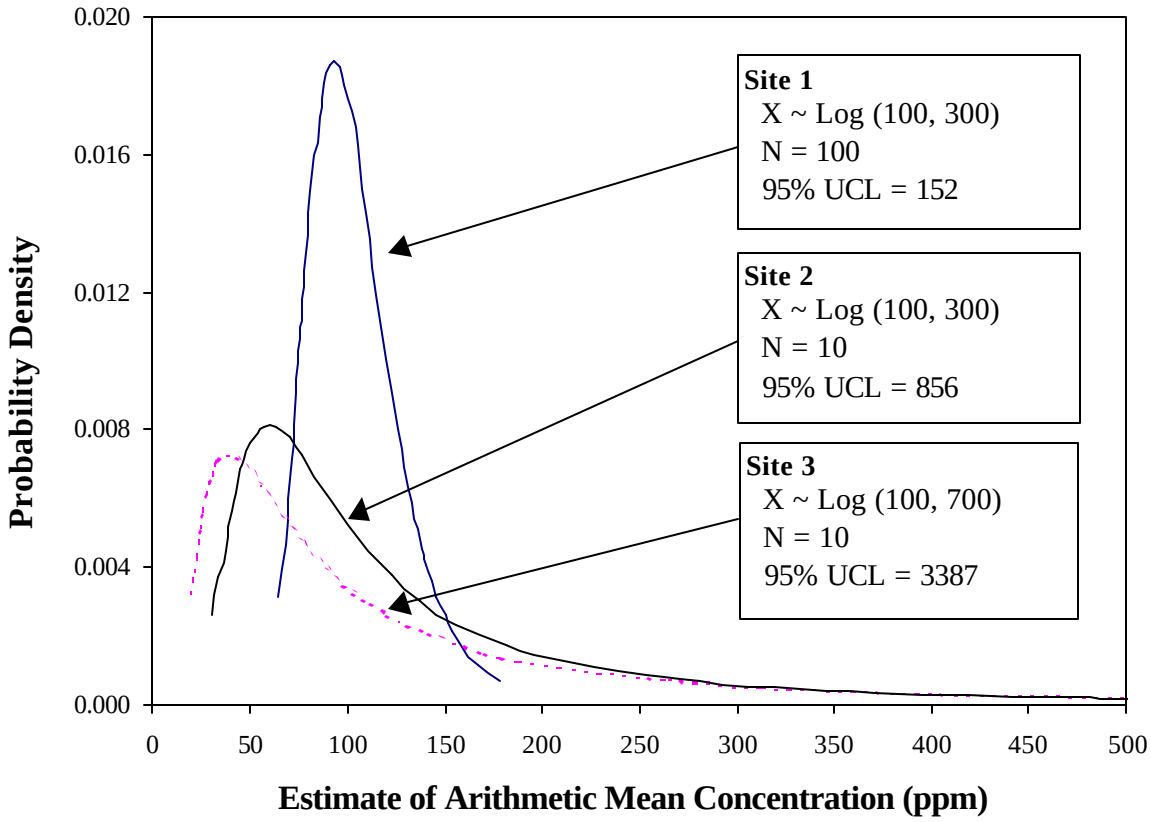
$\bar{y}$	=	arithmetic mean of ln(x)
s_y	=	standard deviation of ln(x)
$H_{0.95}$	=	test statistic for a one-sided 95% confidence limit
n	=	sample size

shown in Equation D-2. In general, the Land method is used to derive point estimates of uncertainty in the mean concentration based on the 95th percentile of the uncertainty distribution for the arithmetic mean.

APPENDIX D (PART 2 OF 2)

ESTIMATING UNCERTAINTY IN THE MEAN CONCENTRATION (CONTINUED)

Figure D-1. Probability distributions of arithmetic mean (AM) concentrations at three sites, at which uncertainty is represented as a function of sample size (n), sample AM, and sample standard deviation (SD). For each site, it is assumed that contaminant concentration (X) follows a lognormal distribution with a sample AM and SD, expressed as [X ~ Log (AM, SD)]. Confidence limits for AM are calculated exactly using an MSDOS program provided by C.E. Land (1997) that employs the following algorithm (Land, 1971, 1975; U.S. EPA, 1992):



In addition to the point estimate calculation, the Land method can be used to calculate the complete distribution of uncertainty by varying alpha (") in the estimate of the H-statistic. Figure D-1 presents distributions for uncertainty that correspond with different lognormal distributions and sample sizes. The complete distribution may be used in a 2-D MCA, whereas the point estimate of the 95% UCL may be used for the concentration term in either 1-D MCA or 2-D MCA.

$$UL_{1-\alpha} = \exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right)$$

where $\bar{y}$ and s_y are the mean and standard deviation of ln(X), and $H_{1-\alpha}$ is the test statistic given an upper one-sided 100(1 - ")% confidence limit and sample size *n*. Note that EPA (1997) demonstrates that use of the Land method will yield wider confidence limits than other methods (e.g., bootstrap).

D.2 BOOTSTRAP METHOD

Bootstrap refers to a method for estimating confidence intervals for a statistic by resampling a data set with replacement to form new data sets (called bootstrap samples) with the same sample size as the original data set. A statistic of interest, such as the arithmetic mean, is calculated for each bootstrap sample. The statistics generated from the bootstrap samples are referred to as replicates.

Many different bootstrap methods have been developed to estimate confidence intervals from bootstrap statistics (see highlight box below). More detail is provided in EPA's *Lognormal Distribution in Environmental Applications* (U.S. EPA, 1997) and the thorough evaluation by Hall (1988). Bootstrap-t is presented in this appendix. In general, confidence intervals are determined from the standard error of the replicates, or from the cumulative distribution function for the replicates. Each bootstrap sample can be obtained using Monte Carlo resampling techniques. The number of bootstrap samples, B, appropriate for developing reliable confidence limits depends on the statistic of interest and the acceptable error in the interval. A minimum of B = 1000 is recommended by Efron and Tibshirani (1993).

EXAMPLES OF BOOTSTRAP METHODS

- standard bootstrap
- bootstrap-t (also known as pivotal method
- percentile
- bias correction approach (BCa)

There are three main advantages of using bootstrap techniques to characterize uncertainty: (1) estimates the standard error of a statistic for which an equation of standard error is either extremely complex or non-existent; (2) estimates the standard error without fitting the sample to a parametric distribution; and (3) techniques are relatively easy to implement on a computer. According to the central limit theorem, the arithmetic means obtained from independent, random samples drawn from the same population will be approximately normally distributed, regardless of the distribution of the sampled population, if the sample size is large (Snedecor and Cochran, 1989). When the assumption of normally-distributed means is valid, confidence intervals (CIs) for the mean may be estimated using the t-statistic. Note that the more familiar z-statistic is appropriate when the variance of the population is known. In

practice, however, small sample sizes can yield inaccurate estimates of confidence intervals for moderately skewed distributions. As *n* increases, the validity of assuming a normal distribution of means increases. The size of *n* required for the assumption of normality to be valid depends on the variability and skewness of the data (U.S. EPA, 1997; Chen, 1995).

Analytical solutions have been developed for estimating CIs for the mean for samples that are obtained from skewed distributions (e.g., Chen, 1995). As discussed above, Land (1975) developed a method appropriate for lognormal distributions.

$$CI = (LCL, UCL) = \left(\hat{\mu} - t_b^{1-0.5\alpha} \cdot \hat{s}_{e,b}, \hat{\mu} - t_b^{0.5\alpha} \cdot \hat{s}_{e,b} \right)$$

where,

$$t_b = \left(\frac{\hat{\mu}_b^* - \hat{\mu}}{\hat{s}_{e,b}^*} \right)$$

$\hat{\mu}_b$ = mean of bootstrap sample
$\hat{\mu}$ = mean of sample data
t_b = estimate of t for the b^{th} bootstrap
$\hat{s}_{e,b}^*$ = estimated standard error of the mean for the bth bootstrap sample

EXAMPLE D-1. COMPARISONS OF METHODS FOR QUANTIFYING CONFIDENCE LIMITS FOR THE ARITHMETIC MEAN OF RIGHT-SKEWED DISTRIBUTIONS.

The following example illustrates the effect that skewed distributions have on estimates of confidence intervals for the arithmetic mean. Assume Tables D-1 and D-2 summarize the contaminant concentrations (ppm) that were detected in surface sediments at three different sites.

Table D-1. Sample Data

Sample A	Sample B	Sample C
34.8	2.2	2.9
38.0	2.4	4.2
58.1	2.5	15.0
71.1	44.5	18.2
83.8	50.0	24.6
102.2	112.3	29.1
102.2	152.7	270.8
109.8	1786.5	2280.6
109.9	1849.4	250000.0
192.4	10917.8	500000.0

Table D-2. Sample Statistics

Statistic	Sample A	Sample B	Sample C
n	10	10	10
$\bar{x}$	90	1492	75265
s	46	3393	168608
GSD[1]	1.612	3.853	3.818
g [2]	1.1	2.9	2.3
percentile[3]	48.3th	75.7th	81.0st

[1] Geometric standard deviation, assuming sample data can be fit to a lognormal distribution.

[2] *g* = coefficient of skewness.

[3] Distribution percentile that corresponds to $\bar{x}$.

Sample Statistics

Summary statistics for three hypothetical samples are provided in Table D-2. Each sample has the same number of measurements (n = 10) and can be described by a probability distribution that is skewed to the right (i.e., positive coefficient of skewness). Sample A has no extreme values in the tail, resulting in a low coefficient of skewness (g = 1.1). Sample B has a single extreme value (10917.8), which tends to inflate the arithmetic mean, standard deviation, and coefficient of skewness. Similarly, sample C has two measurements with extremely high concentrations (i.e., two orders of magnitude greater than other measurements), yielding the highest sample mean and standard deviation. Although the coefficient of skewness for sample C is lower than that of sample B (i.e., 2.3 vs. 2.9), the distribution percentile corresponding to the mean is the highest (i.e., 81st percentile).

Confidence Intervals for the Mean

Two approaches have been used to estimate confidence intervals for the mean of positively skewed distributions. First, one may assume that the sample data are described by a continuous probability distribution, such as the lognormal distribution (see Chapter 3 for guidance on selecting and fitting distributions). U.S. EPA (1992) provides guidance on estimating the confidence interval (CI) for the mean of a lognormal distribution using the Land method (Land, 1971; 1975; 1997). In general, CI is sensitive to the sample size (n) and variance (s^2); CI tends to increase with decreasing sample size and

increasing variance. A second approach, which encompasses a group of techniques referred to as bootstrap methods, requires no assumption regarding the probability distribution for the sample. U.S. EPA (1997) provides guidance on estimating CI using various bootstrap methods. As with the Land method, CIs estimated using bootstrap methods are sensitive to the sample size (n) and variance (s^2) of the sample.

Table D-3 summarizes the 90% confidence intervals for the arithmetic mean of the three samples using the four (4) different bootstrap methods and the Land method. Detailed descriptions of the bootstrap approaches are provided in U.S. EPA (1997) and Efron and Tibshirani (1993). In addition, results of the central limit theorem (CLT) are also provided. The CLT states that as the sample size increases, the distribution of $\bar{x}$ approaches a normal distribution, no matter what the underlying population distribution may be.

Table D-3. 90% Confidence Intervals for the Mean.[1]

Estimation Method	Sample A ($\bar{x}$ = 90)		Sample B ($\bar{x}$ = 1492)		Sample C ($\bar{x}$ = 75265)	
	Lower	Upper	Lower	Upper	Lower	Upper
1. Central Limit Theorem (CLT)	64	117	- 475	3459	- 22,474	173,003
2. Standard Bootstrap[2]	65	115	- 373	3357	- 17,780	168,309
3. Percentile Bootstrap[2]	69	114	218	3488	270	175,033
4. BCa Bootstrap[2]	71	118	405	4752	25,237	225,034
5. Bootstrap-t[2]	69	121	245	8883	7554	17.0×10^6
6. Land[3]	108	3245	2738	53.9×10^9	6406	94.1×10^9

[1] 90% CI estimated from 1-sided lower and upper 95% confidence limits.
[2] Bootstrap estimates are based on B = 10,000 estimates of $\bar{x}$.
[3] Estimated using Land's MS DOS program (Land, 1997) for exact CI for mean of lognormal distributions.

For sample A, which is only moderately right-skewed, the CLT and four bootstrap estimates yield approximately the same 90% CI, all of which contain the sample mean (90) and the population mean used to generate the original sample (i.e., a lognormal distribution with μ = 100 and σ = 50). Land's method yields a wider CI that does not contain the sample mean.

For sample B, which contains one extreme value and the highest coefficient of skew (g = 2.9), the 90% CI is wider for each method. Again, the CLT and bootstrap methods each contain the sample mean, while the Land method yields a 95% lower confidence limit (LCL) that exceeds the sample mean (i.e., 2738 > 1492). Note that the CLT and standard bootstrap methods also yield a 95% LCL less than 0, which is not uncommon for skewed distributions. Of the four bootstrap methods, the bootstrap-t method tends to yield the widest CI.

For sample C, which contains two extreme values and a high coefficient of skew, the 90% CI is wider still. Each method yields a 95% UCL that is more than two times greater than the sample mean. Both

the bootstrap-t and Land methods are especially sensitive to extreme values, yielding 95% UCL values that are three and six orders of magnitude greater than the sample mean, respectively.

These examples illustrate how, for small sample sizes, estimates of confidence intervals for the mean are sensitive to extreme values in right-skewed distributions. This is not unique to the bootstrap approaches.

REFERENCES FOR APPENDIX D

Chen, L. 1995. Testing the mean skewed distributions. *J. Amer. Stat. Assoc*. 90(430):767-772.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY.

Hall, P. 1988. Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics*. 16(3): 927-853.

Land, C.E. 1971. Confidence intervals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics*. 42: 1197-1205.

Land, C.E. 1975. *Tables of confidence limits for linear functions of the normal mean and variance*. In: Selected Tables in Mathematical Statistics, vol. 3. American Mathematical Society, Providence, RI.

Land, C.E. 1997. *MSDOS program to calculate exact confidence limits for the mean of lognormal distributions*. Personal communication.

Snedecor, G.W. and W.G. Cochran. 1989. Statistical Methods. *Iowa State Univ. Press, Ames, Iowa.*

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. Office of Solid Waste and Emergency Response, Washington, DC. OSWER Directive 9285.7-081.

U.S. EPA. 1997. *The Lognormal Distribution in Environmental Applications*. Office of Research and Development, and Office of Solid Waste and Emergency Response, Washington, DC. EPA/600/R-97/006.

APPENDIX E (PART 1 OF 2)

ADVANCED MODELING APPROACHES FOR CHARACTERIZING VARIABILITY AND UNCERTAINTY

E.0 INTRODUCTION

This appendix briefly describes the following advanced modeling approaches that can be used in PRA to characterize variability and uncertainty: 2-dimensional (2-D) MCA, microexposure event analysis (MEE), geospatial statistics, and Bayesian analysis. The application of many of these approaches will require access to expertise in specialized areas of statistics and, in some cases, specialized or even custom-designed computer software. The intent here is to introduce some of the basic concepts and terminology, as well as to provide references where the reader can find more exhaustive coverage of these topics.

E.1 EXPRESSING VARIABILITY AND UNCERTAINTY SIMULTANEOUSLY

A Monte Carlo analysis that characterizes either uncertainty or variability in each input variable (see Example in Chapter 1) can be described as a one-dimensional analysis (1-D MCA). A two-dimensional Monte Carlo analysis (2-D MCA) is a term used to describe a model that simulates both uncertainty and variability in one or more input variables. All probability distributions that are used to describe variability in a PRA model have a certain degree of associated uncertainty. For example, suppose variability in soil concentration (ppm) is estimated using a normal PDF defined by a mean (μ_{soil} = 5) and standard deviation (σ_{soil} = 1), and subjectively truncated (min, max) at (0, 50). Uncertainty in the parameter estimates can be represented in a PRA model by assuming both parameters are also random variables. To illustrate this concept, assume normal PDFs for *uncertainty* can be specified for both parameters. Uncertainty in the mean is described by the normal PDF with parameters (μ_{mean}= 5, σ_{mean} = 0.5); similarly, uncertainty in the standard deviation is described by the normal PDF with parameters (μ_{SD} = 1, σ_{SD} = 0.5). Model variables are represented in this manner when there is a compelling reason to believe that a unique probability distribution does not adequately describe one's knowledge of each variable in the model. A variable described in this way is called a second order random variable. Figure E-1 (Panel A) shows a collection of n = 20 cumulative probability distributions (CDFs), each curve representing a unique set of (mean, SD) parameter estimates for the normal PDF for variability. Panel B shows the 90% *confidence interval*[1] based on 2500 simulated CDFs. The 95% lower and upper bounds correspond to the distribution of 5th percentiles and 95th percentiles, respectively (i.e., CDF for 2500 5th percentiles and CDF for 2500 95th percentiles). The 90% CI for the 50th percentile is (3.4, 6.7).

[1]Note that the term "credible interval" may be more appropriate than "confidence interval" given that the range is based on subjective as well as statistical considerations. Brattin, Barry, and Chiu (1996) provide additional examples of uncertain PDFs that illustrate this concept.

1

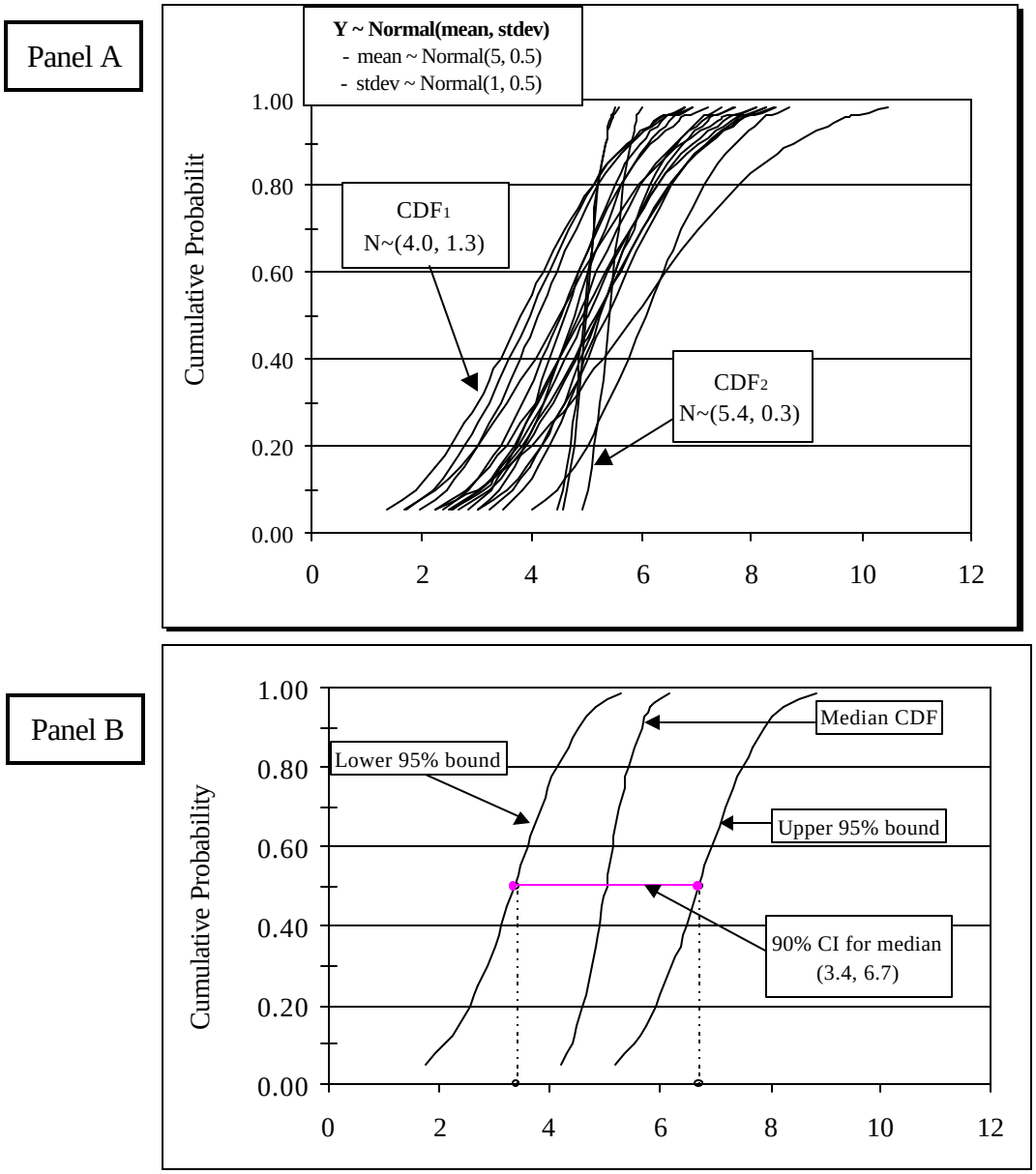


Figure E-1. **Panel A** shows a family of 20 CDFs for a hypothetical random variable, *Y* (e.g., concentration in units of ppm), characterized by a normal PDF where both the mean and SD are also random variables representing uncertainty in the parameter estimates: Mean~ Normal(5, 0.5), SD~ Normal(1, 0.5). Each CDF represents a single simulation of n = 2500 iterations using a unique set of parameters. For example, CDF_1 represents N~(4.0, 1.3) while CDF_2 represents N~(5.4, 0.3). **Panel B** shows the "90% credible interval" for the CDF based on 2500 *simulations*, each simulation using n = 2500 iterations (i.e., a 2-D MCA with 2500 outer loop iterations and 2500 inner loop iterations). Lower, median, and upper bounds represent the simulated 5th, 50th, and 95th percentiles, respectively. The 90% confidence interval for the estimate of the 50th percentile is: {3.4, 6.7}.

DEFINITIONS FOR APPENDIX E

Bayesian statistics - A specialized branch of statistics that views the probability of an event occurring as the degree of belief or confidence in that occurrence.

Geospatial statistics - A specialized branch of statistics that explicitly takes into account the georeferenced context of data and the information (i.e., attributes) it contains.

Frequentist - A term referring to classical statistics in which the probability of an event occurring is defined as the frequency of occurrence measured in an observed series of repeated trials.

Image Analysis - A technique in geostatistics used to restore a degraded image or interpret images that have been contaminated by noise or possibly some nonlinear transformation.

Kriging - A geostatistical method of spatial statistics for predicting values at unobserved locations.

Likelihood Function - A Bayesian term referring to a probability distribution expressing the probability of observing a piece of new information given that a particular prior belief is true.

Location tag - The spatial coordinates of a sampling location (e.g., longitude, latitude).

Microexposure Event Analysis (MEE) - An approach to modeling exposure in which long-term exposure of an individual is simulated as the sum of separate short-term exposure events.

Point Pattern Analysis - A technique in geostatistics of restricting the analysis to location information, ignoring attribute information; addresses two location problems: 1) describing points according to spacing, and 2) describing points according to density.

Posterior Distribution - A Bayesian term referring to a probability distribution that has been updated with new information.

Prior distribution - A Bayesian term referring to the hypothesized, expected, or calculated probability distribution for an event prior to the collection of new information.

Spatial Autocorrelation - The tendency of data from locations that are relatively close together to be geographically correlated.

Thiessen (Voronoi) polygon analysis - A method of spatial statistics in which an area is subdivided into subregions, or polygons, in order to predict values at unobserved locations.

Time Step - A modeling term used to describe the time interval within which variable values do not change.

Two-Dimensional (2-D) Monte Carlo Analysis (MCA) - Separate representation of variability and uncertainty in an MCA, usually accomplished using nested computation loops.

In the example shown in Figure E-1, the mean and standard deviation for soil concentration were allowed to vary independently. Thus, a distribution could be defined by a combination of a low mean and a high standard deviation or a high mean and low standard deviation, or any other combination in between. The assumption of independence of variable parameters may not be valid in all cases. It may be unreasonable to assume that a high mean soil concentration would occur with a low standard deviation. An alternative assumption would be that the standard deviation of the mean is a constant proportion of the mean (i.e., a constant coefficient of variation). Correlations between parameters should be considered in the design of the PRA. One approach that is especially useful for characterizing relationships between the slope and

intercept of a simple linear regression is to specify the bivariate normal distribution for the parameter estimates.

E.2.1 TWO-DIMENSIONAL MONTE CARLO ANALYSIS (2-D MCA)

Two-dimensional Monte Carlo analysis (2-D MCA) is an approach for computing risk (or hazard) when combining distributions that represent variability and uncertainty. In 2-D MCA, distributions representing variability and uncertainty are sampled using nested computational loops (Figure E-2). The inner loop simulates variability by repeatedly sampling values for each variable from their defined probability distributions. With each circuit of the outer loop, new parameter values for each variable are selected, and the inner loop sampling is repeated. The result is a collection of inner loop simulations, one for each parameter value selected. If the inner loop samples 5000 times, and the outer loop samples 1000 times, then each variable is sampled 5,000,000 times and 1000 simulated probability distributions of risk are generated from the PRA model. These probability distributions can be analyzed to estimate the distributions for specific risk estimates. For example, confidence limits on the estimate of specific risk percentiles can be simulated using 2-D MCA (Figure E-3).

E.3 MICROEXPOSURE EVENT ANALYSIS

The standard intake equation generally used in Superfund site risk assessments expresses the time-averaged intake as the products of the time-averaged, body mass adjusted, exposure concentration, medium intake rate and exposure duration (Equation E-1). If the risk assessment is directed at assessing life-time risk to humans, the averaging time used in Equation E-1 would generally be 70 years (i.e., estimated average human lifetime), and the calculated chemical intake would generally represent the life-time average daily intake (LADD). Where information is available to characterize variability on a smaller time scale than life-time, an alternative expression of intake that accommodates such variability may be desirable.

Standard Time-Averaging

$$INTAKE = \frac{C \cdot IR \cdot EF \cdot ED}{BW \cdot AT}$$ Equation E-1

Microexposure Event Modeling

$$NTAKE = \frac{1}{AT} \sum_{j=1}^{ED} \frac{1}{BW_j} \sum_{i=1}^{Events_j} C_{ij} \cdot IR$$ Equation E-2

C = Concentration; i = exposure event; j = year of life
IR = Intake Rate
EF = Exposure Frequency
ED = Exposure Duration
BW = Body Weight
AT = Averaging Time

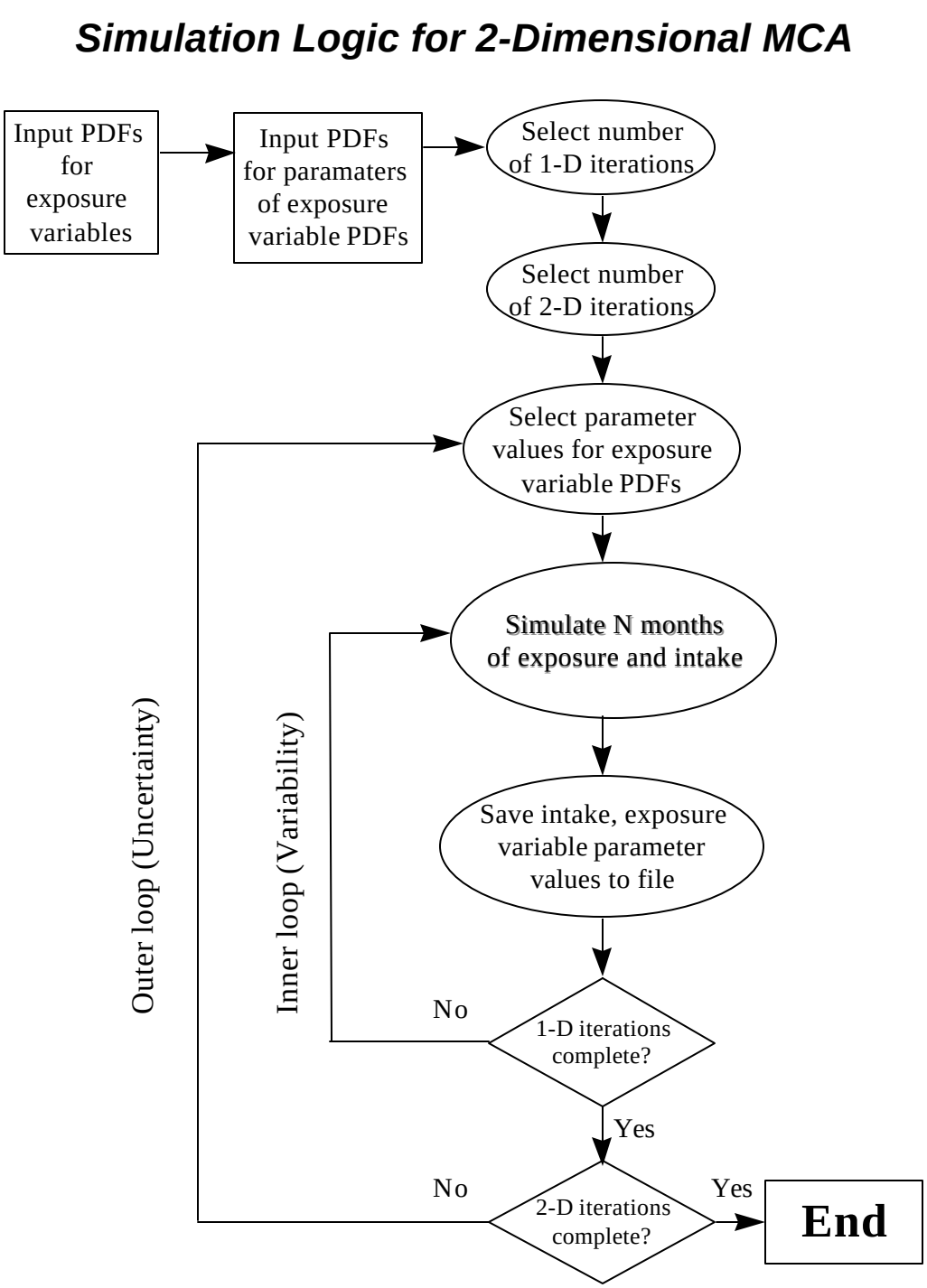


Figure E-2. Diagram showing of a 2-D Monte Carlo model in which the variability and uncertainty dimensions are computed in nested loops. In this example, values for exposure variables in the inner loop represent monthly averages.

1

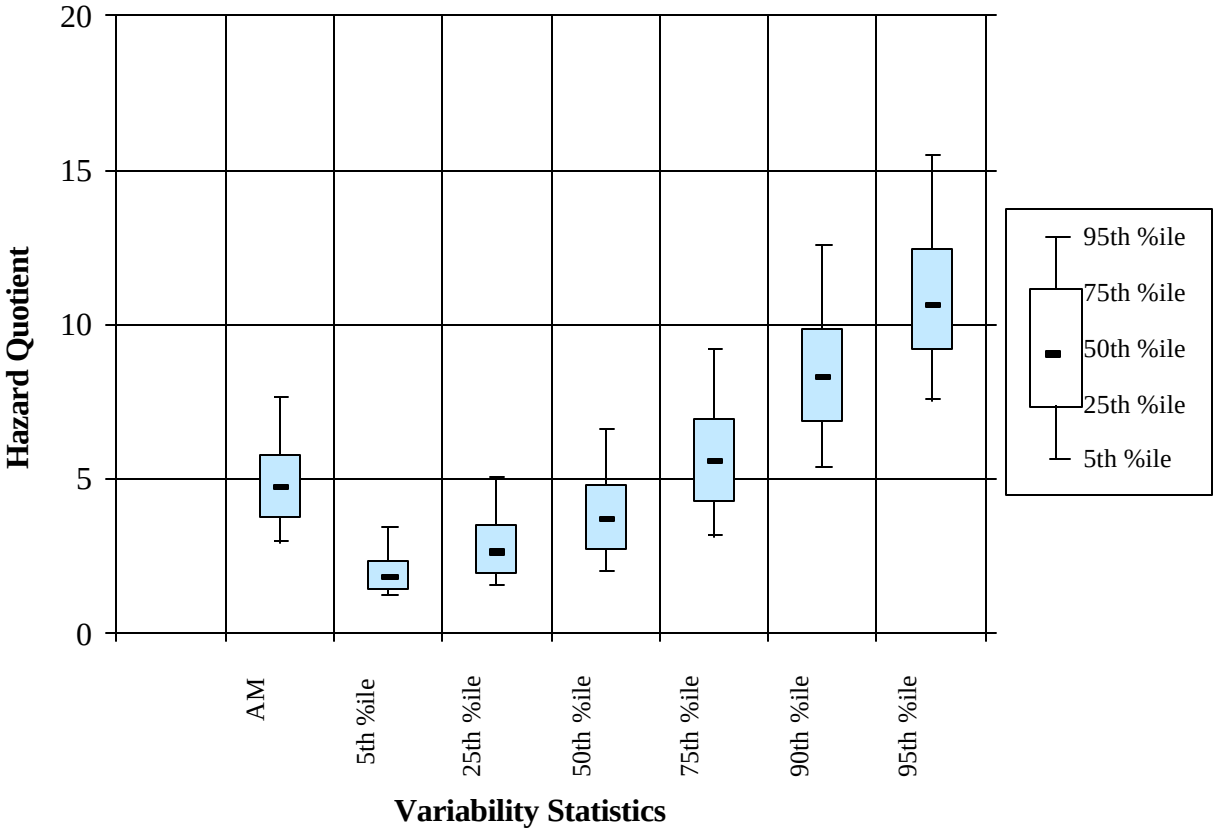
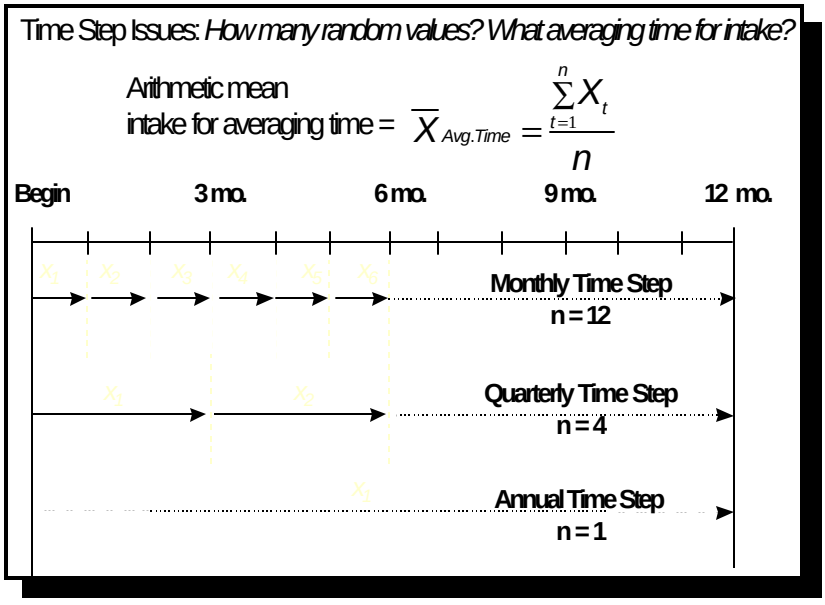


Figure E-3. Output from a 2-D MCA showing the estimated mean HQ and the 90% confidence interval for selected percentiles of the HQ distribution. The 95th %ile HQ would be the RME risk estimate. The simulation suggests that there is a 95% probability that the RME HQ (95th percentile) is below 16.

Concentrations in various environmental media can be expected to vary over time. For example, wind erosion may change chemical concentrations in surface soil. Leaching may change concentrations in both subsurface soil and groundwater. The change in the concentration term is most readily apparent when considering harvesting anglers. If an angler consumes a large amount of fish from a single location (e.g., a specific lake, pond or river), then the average chemical concentration in the fish consumed by that angler can be expected to be similar to the average of the chemical concentration of fish in the population. However, if an angler consumes fish only occasionally, or harvests fish from different locations, there will be considerably more uncertainty in the concentration term. In addition, a harvesting angler may consume varying amounts of fish over the period of the exposure duration due to changing tastes, changes in the fish population size or other factors.

Daily activity patterns, food intake, soil ingestion and other behavioral factors are measured in a time period of less than a year. The extrapolation of these short term results to the chronic exposure situation is a source of uncertainty. Exposure events are real but unknowable whereas data regarding the nature and magnitude of these events is known but its application to a real world situation is uncertain. Microexposure event analysis (MEE) attempts to reduce this uncertainty (Price et al., 1996). MEE modeling provides an alternative to the standard time-averaging approach represented by Equation E-1. In the MEE approach, long term intake is viewed as the sum of individual exposure events (Equation E-2). Implementing the MEE approach in a PRA requires dividing the exposure duration into short epochs, or time steps, within which the values assigned to exposure variables remain constant, but are allowed to vary from one time-step to the next. In a PRA model, exposure variables are adjusted at each time step by selecting values from the probability distributions representing each variable (Figure E-4). Discussion of the implementation of MEE analysis to risk assessment and its merits and limits can be found in Buck et al. (1997), Price et al. (1996), Slob (1996), and Wallace et al. (1994).



In MEE modeling, the time step becomes an important variable, with associated uncertainty. The time step should be selected based on information available to describe how exposures change over time. For example, a model of a moving plume of solvents in groundwater might suggest that chemical concentrations in a given location are dropping by between 16 and 25 percent quarterly. Several rounds of sampling may support this

prediction. This rapid decline in concentrations suggests that an appropriate time step might be one quarter (i.e., three months).

On the other hand, where risk is being assessed for metals, dioxin or PAHs in soil, the concentrations might be expected to change much more slowly, if at all, and the basis of the time step might be the increase in age and corresponding changes in behavior of the receptor. The time step may be global; that is, one time step may apply to all variables in the model. In this case, the same number of random values would be selected for each exposure variable in a Monte Carlo simulation. A more complex model may use different time steps for different variables, requiring some probability distributions to be sampled more often than others. The selection of a value for a time step implies that the value represents the average value for that variable during the time step.

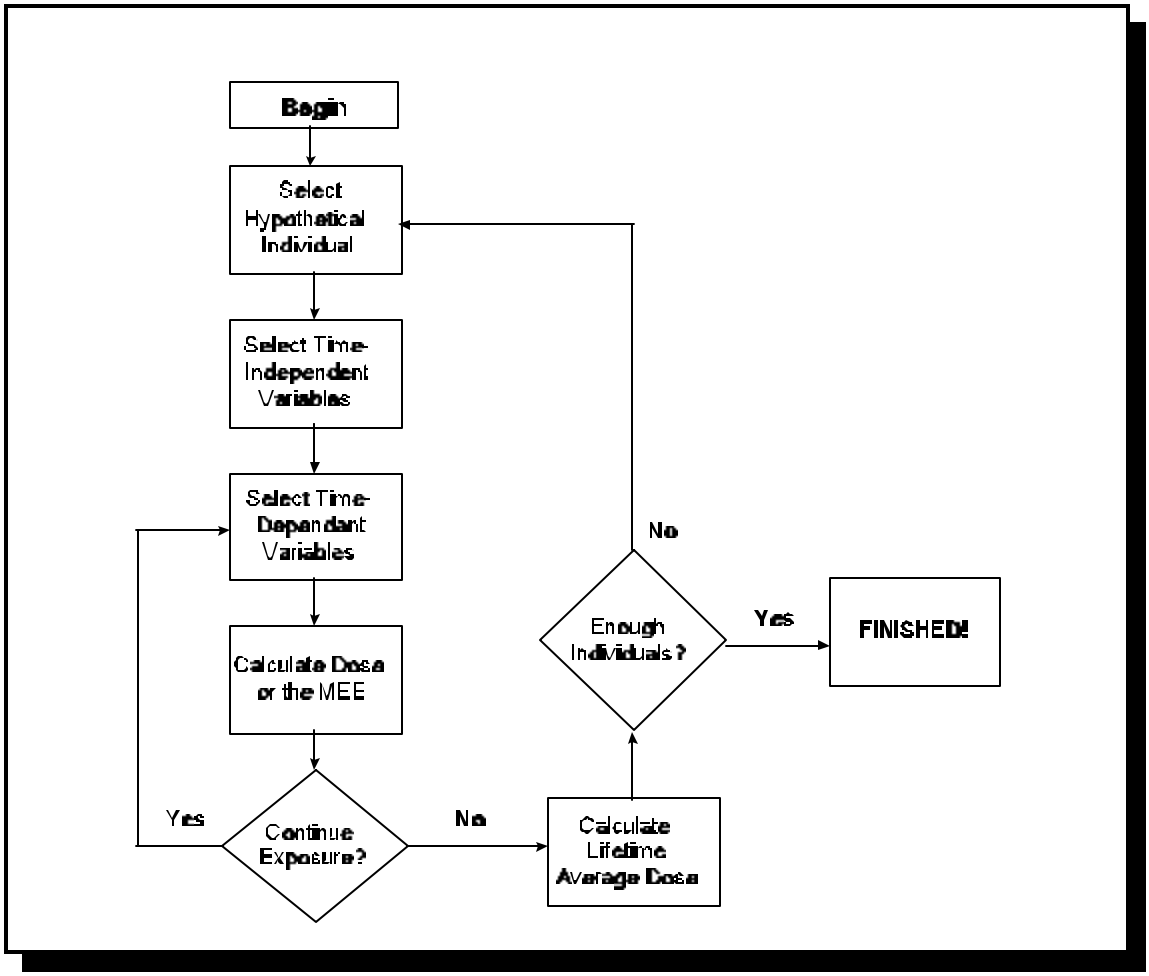


Figure E-4. Flowchart showing general approach for Microexposure Event (MEE) analysis.

Two important issues related to time-step need to be considered in implementing the MEE approach in PRA models. The first is the relationship between the length of the time step and the number of times random values are generated from a defined probability distribution. As the time step decreases, more time steps are needed to simulate exposures over a specified duration. For example, given a time step of one year and an exposure duration of 30 years, each random variable will be sampled 30 times (once per year); for a time step of one month and an exposure duration of 30 years, each random variable would be sampled 360 times (i.e., 12 months/year x 30 years). The Central Limit Theorem indicates that as *n* increases, the distribution of sample means is approximately normal, and the standard deviation of the sample distribution is inversely proportional to the square root of *n*. Thus a highly skewed input distribution (e.g., lognormal) may tend to become less skewed with increasing *n* (Figure E-5). A biased estimate of the RME risk in a PRA model may result if an inappropriately small or large time step is used in the model. This emphasizes the importance of having an empirical basis for selecting the time step and of exploring the time step as a variable in a sensitivity analysis of the model.

The second issue related to the time step concerns temporal correlations. Is it reasonable to assume that random values selected for consecutive time steps are completely independent? For example, consider body weight. The body weights of an individual measured at different times would be expected to show positive temporal autocorrelation; that is, body weight is likely to be similar (but not constant) from one time step to the next. For example, if an individual weighs 60 kg during one month, it is unlikely that they will weigh 80 kg the next month. If this scenario is accepted, then body weight should not be allowed to vary independently from one monthly time step to the next in the model. At shorter time steps, temporal correlation becomes more likely as a result of temporal autocorrelation. For example, one can expect a higher correlation between body weights on an individual measured on two successive days (one-day time step) than between weights measured at the midpoint of two successive years. Approaches to simulating temporal correlations in probabilistic models might include fixing an individual within a percentile range of a distribution (e.g., randomly assigned quartile) or using randomly assigned fluctuations (e.g., $BW_t = BW_{t-1} \pm x$).

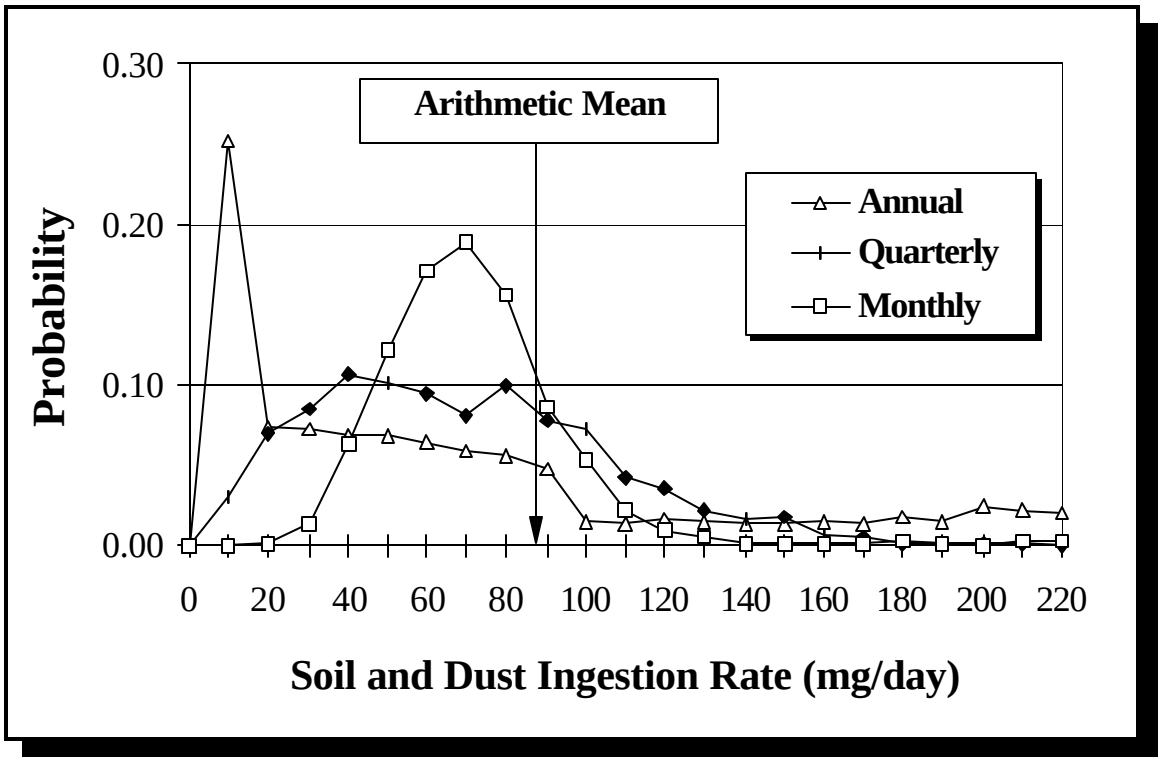


Figure E-5. Hypothetical example showing the effect of model time step on the probability distribution for soil and dust ingestion rate in children over a 1-year period. Number of samples (*n*) needed to simulate exposures: Annual (1), Quarterly (4), Monthly (12).

E.4 GEOSPATIAL STATISTICS

Spatial statistics is a specialized branch of statistics, falling under the heading of multivariate statistics, that explicitly takes into account the georeferenced or locational tagged context of data. All environmental samples collected at Superfund sites have such geocoding. By acknowledging the geography of site chemicals, information about the spatial distribution of contamination can be incorporated into an exposure assessment. This can lead to a more accurate estimate of the confidence limits for the arithmetic mean concentration. Geospatial statistics quantifies the spatial autocorrelation of sample measurements and allows for the exploration of the spatial distribution of exposure and risk using techniques of map generalization. By recording locational tags for each sample, information about spatial patterns within an exposure unit can be exploited to estimate both pre- and post-remediation exposure and risk.

EXHIBIT E-1

POSITIVE SPATIAL AUTOCORRELATION

- locations with a high value of Y tend to be surrounded by nearby high values of Y
- locations with a medium value of Y tend to be surrounded by nearby medium values of Y
- locations with a low value of Y tend to be surrounded by nearby low values of Y

Several important risk assessment issues are closely linked to geospatial statistics, as described in Exhibit E-2. Geospatial statistics comprises:

- *spatial autoregression,*
- *geostatistics,*
- *point pattern analysis, and*
- *image analysis*

The first three of these subjects can contribute to spatial statistical support of site risk assessments. The key concept linking all three is spatial autocorrelation, which refers to covariation among samples for a single chemical, or the tendency of data from locations that are relatively close together to be geographically correlated. By analogy, classical statistics treats soil samples as though they are balls, each having a battery of attributes, that can be placed into an urn for statistical analysis; geospatial statistics treats soil samples as though they are clusters of grapes, with the branchy stems representing locational

EXHIBIT E-2

EXAMPLES OF RISK ASSESSMENT ISSUES LINKED TO GEOSPATIAL STATISTICS

- Sampling tends to disproportionately represent "hot spots" (i.e., a relatively large portion of a data set with a small sample size (n) tends to be concentrated at "hot spots").
- The upper confidence limit (UCL) for the arithmetic mean exposure concentration (e.g., chemical concentrations in soil) depends on the sample size.
- Additional sampling may be needed, especially to better define the spatial patterns or the extent of contamination.
- There is uncertainty about locations not sampled at a site, as well as uncertainty regarding the representativeness of neighboring samples in nearby exposure units.

tags. Concentrations located on the same "branch" will be more strongly correlated than concentrations on different branches.

E.4.1 CORRELATION AND SPATIAL AUTOCORRELATION

Several simple bivariate statistical approaches (also see Chapter 3) may be used to introduce the concept of spatial autocorrelation. Consider two variables, X and Y. For positive correlation there is a tendency for high values of X to be paired with the high values of Y; medium values of X to be with the medium values of Y, and low values of X with the low values of Y. The tendency is in the opposite direction for negative correlation; high values of X tend to be paired with low values of Y, and so on. Spatial autocorrelation, which virtually always is positive, directly parallels these definitions, but is written in terms of a single variable as shown in Exhibit E-1.

Just as the bivariate relationship between X and Y can be portrayed by a scatterplot (Y versus X), the spatial autocorrelation relationship can be portrayed with a Moran scatterplot (sum or average of nearby values of Y versus Y); this is most effective when Y has been converted to z-scores. As shown in Figure E-6, scatterplots can be used to illustrate some important issues related to sample size.

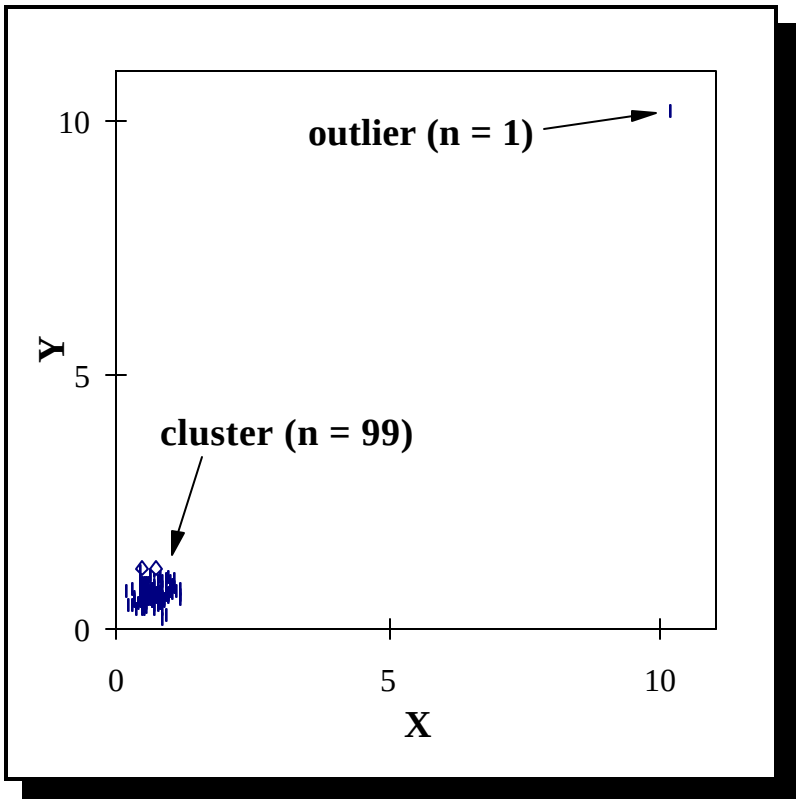


Figure E-6. Effect of an outlier on measured correlation: r = 0.956 with outlier (n = 100), whereas r = 0.086 excluding outlier (n = 99 clustered points).

If no soil samples were collected at a site (n = 0), there is no information about the chemical concentrations in soil, and any guess may be considered an estimate. However, if the chemical concentration of a single sample (n = 1) is measured, some information is obtained that partly restricts this estimate. As each additional independent sample is taken, more information is obtained, and the restriction on the estimate becomes more binding. If the same location is selected repeatedly for sampling, then the repeated measures, which may vary through time, will tend to be highly positively correlated; part of the information obtained from each sample is the same, and should not be counted more than once in estimating the site-wide soil concentration. Similarly, if immediately adjacent locations are sampled, the measures will often tend to be highly positively correlated (spatial autocorrelation). Once the first sample is taken, each additional sample provides only a fractional increment of new information about the site in its entirety.

E.4.2 Effective Sample Size (n*) and Degrees of Freedom

Repeated measures can result in data clustering, which can be illustrated in a scatter diagram. Because two points determine a straight line, if (*n-1*) points cluster together on a scatter diagram while a single additional point occurs far away from this cluster (i.e., an outlier), then the resulting bivariate correlation will be very high (see Figure E-6). This situation alludes to the notion of effective sample size (n*): the n* is no longer equal to the number of observations (n), but rather is dramatically reduced by the presence of inter-observational correlation. For the example shown in Figure E-6, n* is slightly greater than 2 rather than 100 (i.e., n).

Spatial autocorrelation plays an analogous role in georeferenced data. If a sampling network is arranged as a 25-by-25 square grid (one sample point per grid cell), and superimposed over a large site so that a very large distance separates nearby sample locations, then essentially zero spatial autocorrelation should be present in the geographic distribution of the concentrations of any given chemical. Concentrations will appear to be haphazard across the site, rendering the effective sample size as n*=625. If the distance between nearby locations on the sampling mesh is decreased so that the spatial correlation is only r=0.050, then the effective sample size decreases to n*=514. The effect of reducing the inter-sample distance on spatial autocorrelation and n* for a 25-by-25 grid is shown in Exhibit E-3. If r increases to 1, then n* reduces to 1. Therefore, obtaining a measure of latent spatial autocorrelation is essential to estimating n*; this in turn is critical to determining confidence limits for estimates of mean concentrations, which are sensitive to sample size. The upper confidence limit (UCL) for the mean will be biased *only* when very high levels of spatial autocorrelation are present; this is because the

Exhibit E-3

Effect of Spatial Autocorrelation (r) on Effective Sample Size (*n)**

r	*n**
0.000	625
0.050	514
0.539	64
0.957	3
1.000	1

Student- t statistic used to estimate the UCL (assuming a normal distribution) changes very little as the degrees of freedom (related to sample size) increases above 10; part of the difference between *n* and *n** is offset by an inflation of the variance.

The concept of effective degrees of freedom is important in exposure assessment because high positive spatial autocorrelation can bias the estimate of the UCL concentration if geospatial statistics are not considered. This should be of particular concern when specific locations at a site are intensively sampled (e.g., suspected "hot spots"), and other locations are relatively undersampled. Accordingly, the design of the sampling network itself can be evaluated from the perspective of geospatial statistics in order to ascertain the quality of sample information. The ideal sampling network should provide geographic representativeness, should be roughly uniformly distributed over a site, and is best implemented as a stratified random sampling design; that is, the site is partitioned into geographic strata (e.g., exposure units), and then a random sampling of points is selected within each strata. In practice, sample designs may need to focus on objectives that are in conflict with the above ideals. For example, intense sampling of suspected "hotspots" may be necessary at some sites, at the expense of a more representative spatial coverage of the site. In such cases, several statistical techniques are available for assessing the statistical benefit (in terms of reducing uncertainty) of additional sampling at undersampled locations.

E.4.3 Assessment of Additional Site Sampling

Thiessen Polygons. In addition to calculating nearest neighbor statistics, the adequacy of a sampling network can be assessed by Voronoi (i.e., Thiessen polygon) surface partitioning, a popular approach used in mapping intra-site geographic distributions. This procedures divides a site into a mutually exclusive set of polygons, each polygon containing a single measured concentration. Each polygon has the unique property that any location within the polygon is closer to the polygon's sample location than to any other sample point (Clifford et al., 1995). The concentration measured at the sample point in the polygon is assigned to the entire area of the polygon. The intensity of sample points on a surface can be measured by Equation E-3 mean inverse polygon areas:

$$SI = \frac{1}{m}\sum_{i=1}^{m} A_i^{-1} \qquad \text{Equation E-3}$$

where *SI* is a measure of the sampling intensity, A_i is the area of the i^{th} polygon, and *m* is the number of interior polygons (those not along the edge of the site); $m < n$. The variance of the sampling intensity can be expressed by Equation E-4:

$$SI_{variance} = \frac{1}{m-1}\left[\sum_{i=1}^{m} A_i^{-2} - \frac{1}{m}\left(\sum_{i=1}^{m} A_i^{-1}\right)^2\right] \qquad \text{Equation E-4}$$

If the sampling network is uniform (i.e., polygon areas are equal), the variance will be essentially zero. The variance will increase as the network deviates from uniform. This measure can be used to assess whether or not additional samples will improve the spatial coverage.

- Sampling locations that would yield a dramatic reduction in the variance should be given priority for future sampling efforts.

Thiessen polygons can be used to develop area-weighted estimates of the arithmetic mean concentration ($C_{soil,w}$) according to the following general equation:

$$C_{soil,w} = \sum_{i=1}^{n} C_i \frac{A_i}{A_T} \qquad \text{Equation E-5}$$

where C_i is the concentration in the i^{th} polygon, A_i is the area of the i^{th} polygon in the exposure unit, and A_T is the total area of the exposure unit. The weight for each measurement is essentially the ratio of the area of each polygon to the total area of the site. Clifford et al. (1995) applied this approach to an ecological risk assessment of the burrowing owl with the following simplifying assumptions: habitat range is circular, size of exposure unit is constant (75 ha) although location may vary, and organisms spend equal time in all portions of their habitat. Given these assumptions, a nonparametric Bootstrap method can be used to determine the approximate 95% UCL for the mean concentration (see Appendix D). Using Monte Carlo analysis, $C_{soil,w}$ can be estimated for different locations of the exposure unit according to Equation E-3, and confidence limits can be generated from the multiple Bootstrap estimates. Burmaster and Thompson (1997) demonstrate a similar approach in which the exposure unit (with constant area but random rectangular dimensions) is overlayed on the Theissen polygon surface and 95% UCL for the mean is calculated from the Bootstrap sample.

Linear Regression. Another diagnostic is found in the linear regression literature. The locational tag coordinates (e.g., longitude, latitude) can be converted to z-scores (say z_u and z_v) for the following calculation:

$$Y = \frac{1}{n} + \frac{z_u^2 + z_v^2 - 2r_{u,v} z_u z_v}{(n-1)(1-r_{u,v}^2)} \qquad \text{Equation E-6}$$

where Y is a measure of the sampling network, r_{uv} is the correlation between the coordinate axes, and n is the number of samples. Any sampling location (z_u, z_v) in which $Y > 9/n$ may be considered too isolated in the sampling network. Additional sampling locations should be positioned closer to it to improve the overall coverage of the sampling network.

E.4.4 MAP GENERALIZATION

Another important application of geospatial statistics to risk assessment is that of map generalization, which draws on the subjects of geostatistics and spatial autoregression. Techniques developed for both topics exploit spatial autocorrelation in order to produce a map.

Kriging and Semivariograms. Geostatistics employs kriging, which yields statistical guesses at values of a chemical at unsampled locations based on information obtained from sampled locations. Kriging assumes that the underlying geographic distribution is continuous, evaluates spatial autocorrelation in terms of distance separating sample points, and employs a scatter diagram similar to the Moran scatter plot to portray this relationship (i.e., the semivariogram plot: half the squared difference between measured concentrations for two sampled locations versus distance separating these two locations). The best-fit line to this scatter of points is described by one of about a dozen equations (semivariogram models).

Thiessen Polygons and Spatial Autoregression. Spatial autoregression assumes a discretized surface, uses the Thiessen polygon surface partitioning to construct a Moran scatter plot, and can be used to estimate values at selected points with a regression-type equation. Theoretically, the exponential semivariogram model relates to the conditional autoregressive model, and the Bessel function semivariogram model relates to the simultaneous autoregressive model; in practice, though, the spherical semivariogram model often provides the best description of a semivariogram plot. Regardless of which approach is taken to map generalization, one relevant contribution of these two subjects is the following observation:

- Including positive spatial autocorrelation results in more accurate variance estimates; this in turn yields more accurate estimates of the 95% UCL for the mean concentration.

E.4.5 IMPLEMENTATION ISSUES RELATED TO GEOREFERENCED DATA

Estimation of parameters, for either geostatistical or spatial autoregressive models, cannot be achieved with ordinary least squares (OLS) techniques; nonlinear least squares must be used. While OLS provides unbiased regression coefficients, these estimates are not necessarily sufficient (i.e., they do not summarize all of the information in a sample pertaining to the population), efficient (i.e., the standard errors most often are incorrect), and consistent (i.e., the asymptotic sampling distribution concentration will not be at the parameter value). In other words, OLS essentially uses the wrong degrees of freedom in its calculations, as described in Section E.3.2. Two additional complications of georeferenced data that do not appear in other types of data are (1) spatial autocorrelation might be directional (i.e., directional dependency), and (2) variance might be nonconstant over space as well as over the magnitude of the dependent variable, Y (e.g., chemical concentration). Several statistical approaches, which are beyond the scope of this guidance, are available for analyzing these potential sources of bias in the exposure concentration estimates (Isaaks and Srivastava, 1989; Cressie, 1991; Griffith, 1993).

E.5 EXPERT JUDGMENT AND BAYESIAN ANALYSIS

Up to this point in RAGS III, risk has been characterized as having a population probability distribution with parameters (e.g., mean, standard deviation) that can, theoretically, be estimated from observation. In theory, risk estimates could be derived by repeatedly measuring risk in subsets of the population of interest (e.g., repeated measurements of site-related cancer risk) . The unstated expectation, or goal, is that the PRA model will accurately simulate this *real* risk distribution. This approach derives from a *classical* view of probability. The *classical* or *frequentist* view defines the probability of an event as the frequency with which it occurs in a long sequence of similar trials. From the *frequentist* perspective, the probability of having a flipped coin land *heads-up* is given by the frequency distribution for heads-up derived from repeated similar trials of coin flips. Unfortunately, for real world decisions such as those informed by Superfund risk assessments, it is unclear what a representative population of similar trials would be.

Bayesian View of Probability. A Bayesian perspective on probability allows distributions to be constructed based on the judgement of an expert in the field. The subjectivist or Bayesian view is that the probability of an event occurring is the degree of belief a person has in the occurrence. Probabilities can be assessed by experts using scientific knowledge, judgment, data, past experience, and intuition. Different people may assign different probabilities to an event, and a single individual may assign different probabilities to the same event when considered at different times. The consequence is that probabilities become conditional and the conditions must be explicitly stated (Sivia, 1996; Howson and Urbach, 1989; Ott, 1995; Morgan and Henrion, 1990). These conditional probabilities can, of course, be updated with new information.

Using the coin flip analogy above, a Bayesian perspective might be that, based on experience with coins, assuming that most coins are *fair*, and that a fair coin would be expected to land heads-up half the time, the expected probability of the tossed coin landing heads-up is 0.5. If the outcome of repeated trials was different from the expected, the Bayesian approach would be to update the probability based on the new data. In the coin flip example, both the Bayesian and frequentist approaches will arrive at the same conclusions, because the outcome is amenable to rigorous experimentation. Where the two approaches can be expected to differ is in the assignment of probabilities to events that can not be rigorously measured; for example, the probability of a site-related cancer risk, or the probability of a child ingesting a specific amount of soil.

The subjective judgment of experts is, therefore, an important tool in the Bayesian approach to risk assessment. For example, the input distributions for a probabilistic risk assessment can be based upon the judgement of one or more experts who rely upon estimates from the literature, data from experimental studies, and any other information they consider relevant. Even when formal elicitations of expert opinion are not done, the final selection of the form and parameters of the input distributions usually involves some subjective judgement by the analyst. There is a rich literature about the protocol for conducting expert elicitations and using the results to support decisions (Morgan and Henrion, 1990). Elicitation of expert judgement has been used to obtain distributions for use in risk assessments (Morgan and Henrion, 1990; EPA, 1997; Hora, 1992) and in developing air quality standards (EPA, 1982).

In addition to providing input distributions for probabilistic risk assessments, Bayesian analysis allows the current state of knowledge, expressed as a probability distribution, to be formally combined with new data to reach an updated information state. The distribution expressing the current knowledge is the *prior distribution* and may be the output of a PRA (Figure E-7). An appropriate *likelihood function* for the data must also be formulated. The likelihood function is based upon an understanding of the data gathering process and is used to determine the probability of observing a new set of data given that a particular risk estimate is true.

EXHIBIT E-4

COMPONENTS OF BAYES THEOREM IN PRA

- input probability distributions for exposure (or toxicity) based on available data or expert judgement
- prior probability distribution for risk based on input probability distributions (output from PRA)
- new data
- likelihood function, expressing the probability of observing the new data conditional on prior risk

Once the prior distribution is determined, the new data values are collected, and the likelihood function is assumed, Bayes theorem (Exhibit E-4) provides a systematic procedure for updating the probabilistic assessment of risk. The updated information state is called the *posterior distribution* and reflects the reduction in uncertainty arising from the new information.

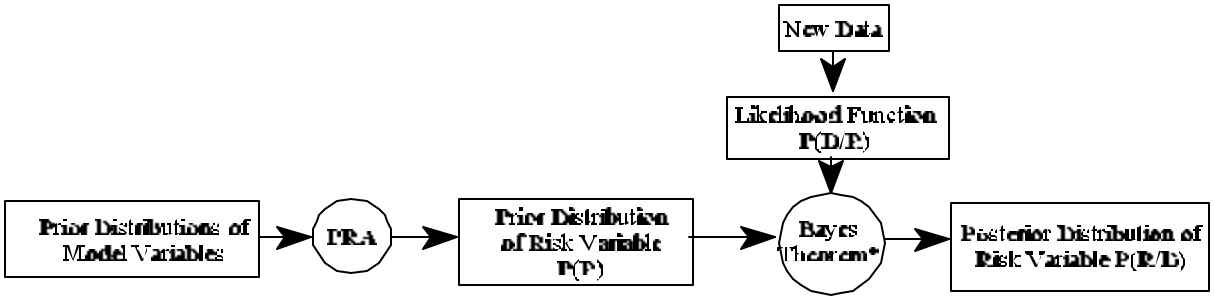


Figure E-7. Conceptual model of Bayesian Monte Carlo Analysis. A PRA simulation yields a prior distribution of risk based on probability distributions for input variables. Given new data for an input variable, and a likelihood function for risk, Bayes Theorem (Eq. E-7) can be used to generate a posterior distribution of risk. The expression P(D/R) refers to a conditional probability, *"the probability of D, given R"*. Conditional probabilities can be thought of as relative frequencies, where R is the information given, and D is the event being computed when a particular value of R occurs.

$$\textit{Bayes Theorem}^{*}: \quad P(R_i/D) = \frac{P(D/R_i)\,P(R_i)}{\sum_{j=1}^{N} P(D/R_j)\,P(R_j)} \qquad \text{Equation E-7}$$

D = new data
R_i = i^{th} risk prediction associated with new data
Rj = j^{th} risk estimate simulated from PRA model
N = number of risk estimates from the PRA model

For example, suppose a model is available to relate soil TCDD concentrations at a site with serum concentrations of TCDD. A probability distribution of soil concentrations is created based upon expert judgement and a limited amount of site specific data. Using the model, the soil concentrations can be associated with a distribution of serum TCDD concentrations (P(R), the prior distribution). New site-specific data (D) are subsequently collected on serum TCDD concentrations in order to reduce uncertainty in the risk estimate. Assume that it is known that serum TCDD concentrations generally follow a lognormal distribution and that the best estimate of the parameters of this distribution come from the prior distribution on serum TCDD. This creates the likelihood function (P(D|R)). Using Bayes Theorem, the new data are used to form a revised distribution of serum TCDD. This is the posterior distribution (P(R|D)).

Bayesian Monte Carlo Analysis. In the past, the use of Bayesian analysis was limited by the degree of mathematical complexity involved. Using Monte Carlo analysis to carry out the PRA, rather than mathematical equations to describe the distributions, allows the calculations to be done much more easily. This variation on traditional Bayesian methods is called Bayesian Monte Carlo analysis (Patwardan and Small, 1992; Dakins et al., 1996). In the TCDD example discussed above and illustrated in Figure E-7, the required calculations are carried out for each of the N iterations of the Monte Carlo analysis (i and j go from 1 to N).

Bayesian Monte Carlo analysis is appropriate in several situations. If a model has been created and a distribution developed using probabilistic risk assessment, new information may be incorporated without the need to repeat the entire analysis. This information could be on one of the uncertain parameters of the model or on the model output variable. Similarly, a generalized risk model with generic parameter distributions may be used for a Superfund risk assessment with the model predictions fine-tuned using data from a particular site of interest. Finally, after a distribution is developed, the amount of uncertainty that exists may be too large for the risk manager to make a decision. In this case, the risk manager should seek out new information that will help refine the analysis and decrease the uncertainty.

Bayesian Monte Carlo analysis can also be combined with techniques from decision analysis to help determine the type and quantity of data that should be collected to reduce uncertainty. Decision analysis is a technique used to help organize and structure the decision maker's thought process and identify a best strategy for action. To determine the appropriate action, one defines the range of possible decisions, evaluates the expected value of the utility or loss function associated with each decision, and selects the decision that maximizes the expected utility or minimizes the expected loss.

- *Decision analysis provides a quantitative approach for evaluating the benefits of including a realistic assessment of uncertainty and the subsequent benefits of reducing this uncertainty.*

Value of Information. Value of information (VOI) analysis involves estimating the value that new information can have to a risk manager before that information is actually obtained (Clemen, 1996). It's a measure of the importance of uncertainty in terms of the expected improvement in a risk management decision that might come from better information. Examples of VOI quantities are the expected value of

including uncertainty (EVIU), the expected value of sample information (EVSI), the expected value of perfect information (EVPI). Calculation of these quantities can be done using mathematical methods, numerical integration (Finkel and Evans, 1987), or Monte Carlo techniques (Dakins, 1999)

Value of information calculations require the specification of either a utility or a loss function. A loss function states the losses associated with making different types of decision errors including both direct monetary costs and losses associated with other consequences. Loss functions take various forms depending on the risk management situation (Morgan and Henrion, 1990).

Expected Value of Including Uncertainty. The expected value of including uncertainty, EVIU, is a measure of the value of carrying out a probabilistic risk assessment. It's the difference between the expected loss of a decision based on a deterministic risk assessment and the expected loss of the decision that considers uncertainty (Figure E-8). If uncertainty in a risk assessment has been estimated using Monte Carlo techniques and a loss function has been specified, the EVIU can be easily calculated. First, the management decision from the deterministic assessment is determined. The loss from making this decision is calculated for each iteration of the Monte Carlo, each time assuming that the risk estimate from that iteration is true. The expected loss is the average of these individual losses. The expected loss for the PRA is determined by calculating the expected loss for a full range of management decisions and selecting the decision with the lowest expected loss. The EVIU is calculated by subtracting the loss associated with the PRA from that associated with the deterministic assessment.

Expected Value of Sample Information. The expected value of sample information is the difference between the expected loss of the decision based on the PRA and the expected loss of the decision from an improved information state. As such, the EVSI is a measure of the value that may result from the collection and use of new information (Figure E-8). Calculation of the EVSI involves a technique called preposterior analysis and is somewhat more complicated.

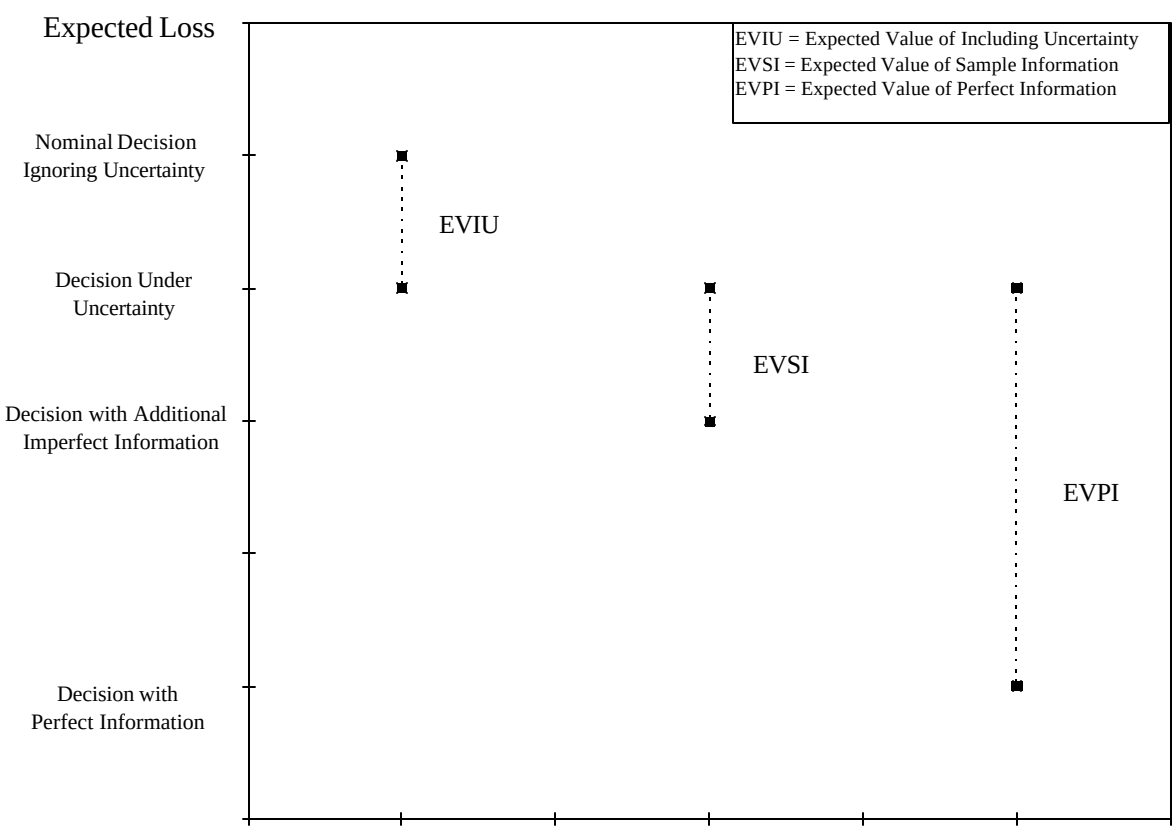
This type of analysis is termed "preposterior" because it involves the possible posterior distributions resulting from potential samples that have not yet been taken. For each replication from the Monte Carlo simulation, the predicted value from the model is used to randomly generate a set of K data points. Each set of data points is then used to calculate the posterior probabilities for the N Monte Carlo simulated values. These posterior probabilities are then used to obtain the optimal answer to the management question at this new level of uncertainty by selecting the decision that minimizes the expected loss over all possible management decisions.

This procedure is repeated for each of the N replications of the Monte Carlo analysis resulting in N posterior distributions, N management decisions, and N associated expected losses. Because each of these outcomes are equally weighted, the expected loss associated with the state of uncertainty expected to exist after the data collection program is carried out is simply the average of the N expected losses. The EVSI is the difference between the expected loss based on the results of the PRA and the expected loss from the updated information state.

Expected Value of Perfect Information. The expected value of perfect information, EVPI, is the difference between the expected loss of the decision based on the results of the PRA and the expected loss of the optimal management decision if all uncertainty were eliminated. In actual application, no research plan or data collection program can completely eliminate uncertainty, only reduce it. The EVPI is an upper bound for the expected value of efforts to reduce uncertainty and so provides the ultimate bound on what should be spent on research and data collection efforts.

Figure E-8. Expected Loss associated with various types of information incorporated into a generic uncertainty analysis.

When a probabilistic risk assessment has been carried out using Monte Carlo techniques, the expected loss associated with perfect information is calculated by determining the expected loss for each iteration of the Monte Carlo, assuming that the correct management decision, if that iteration were true, is made. As always, the expected loss is the average of these losses, and the EVPI is calculated by subtraction.



Uses of Value of Information in Risk Assessment. Value of information analysis has many benefits for risk managers. First, VOI analysis makes the losses associated with decision errors explicit, balances competing probabilities and costs, and helps identify the decision alternative that minimizes the expected loss. VOI analysis can help a decision maker overcome a fear of uncertainty by developing a method to handle it. If the losses associated with making a poor decision are unclear, small uncertainties can take on major importance. Conversely, if the losses associated with different risk management decisions are similar, little additional effort should be expended to continue to consider the alternatives.

APPENDIX E (PART 2 OF 2)

ADVANCED MODELING APPROACHES FOR CHARACTERIZING VARIABILITY AND UNCERTAINTY (CONTINUED)

In addition, VOI analysis helps prioritize spending on research. It provides insights into how resources could be spent to achieve the most cost-effective reduction in uncertainty by identifying which sources of uncertainty should be reduced, what type of data should be obtained, and how much data is needed. Finally, VOI analysis may help decision makers explain the rationale for their decisions to the public and help the public understand the multiple objectives considered in managing risks.

Expected Loss is usually greatest when uncertainty in risk estimates is ignored. For example, by quantifying uncertainty in risk (e.g., 2-D MCA, Bayesian Monte Carlo analysis) a risk manager may determine that the cleanup level associated with the 90th percentile of the risk distribution (rather than the 95th percentile) is adequately protective. Quantifying uncertainty may also result in lower expected loss when more soil remediation is required due to the losses associated with possible under-remediation. The expected loss may be further reduced by collecting additional soil samples, which would presumably reduce uncertainty in estimates of mean exposure point concentrations. The expected loss may be minimized by obtaining "perfect" information (i.e., no uncertainty); however, as shown in the figure, expected value of perfect information spans a wide range of expected loss because the value associated with reducing uncertainty may be tempered by costs associated with additional sampling and analysis. In practice, risk assessors consider this issue when deciding to obtain additional samples for site characterization.

The decision to obtain additional information in order to reduce uncertainty should be made on a site-specific basis, taking into account the potential impact that reducing uncertainty may have on the overall remedial decision. Important questions to consider include: 1) Are the risk estimates sufficiently sensitive to an exposure variable that collecting further data will reduce uncertainty? and 2) Are the confidence limits on the 95th percentile risk estimate sufficiently wide that reducing uncertainty may alter the cleanup goal? An example of decision framework applicable to PRA is presented in Figure E-9. The framework has three tiers. Tier 1 includes the point estimate approach and an assessment of the need for PRA. In Tier 2, the expected value of including uncertainty in the assessment (EVIU) is calculated and, if warranted, a PRA is conducted. In Tier 3, the value of additional information is assessed and Bayes Theorem would be used to incorporate the new information and update probability distributions.

Limitations of These Techniques. Figure E-9 illustrates situations where Bayesian analysis and value of information quantities may not be helpful. For example, if deterministic risk assessment is selected as the appropriate method, these techniques do not apply. Additionally, if site specific data are available that are comprehensive and representative of the population of interest, the results of the Bayesian Monte Carlo analysis will be the same as using the site specific data directly. To be

representative and comprehensive, the data set must be sufficiently large, randomly selected, and represent the full range of variability that exists in the population (e.g., temporal, spatial, inter-individual). However, data sets are rarely perfect, often too small, suffer from relatively high sampling and/or measurement errors, or don't represent the entire population variability over time, space, age, gender, or other important variables. If the data cannot be assumed to describe the population distribution sufficiently well, then probabilistic risk assessment will help to more fully develop the entire range of the population distribution and the Bayesian Monte Carlo analysis will act to refine the model estimates.

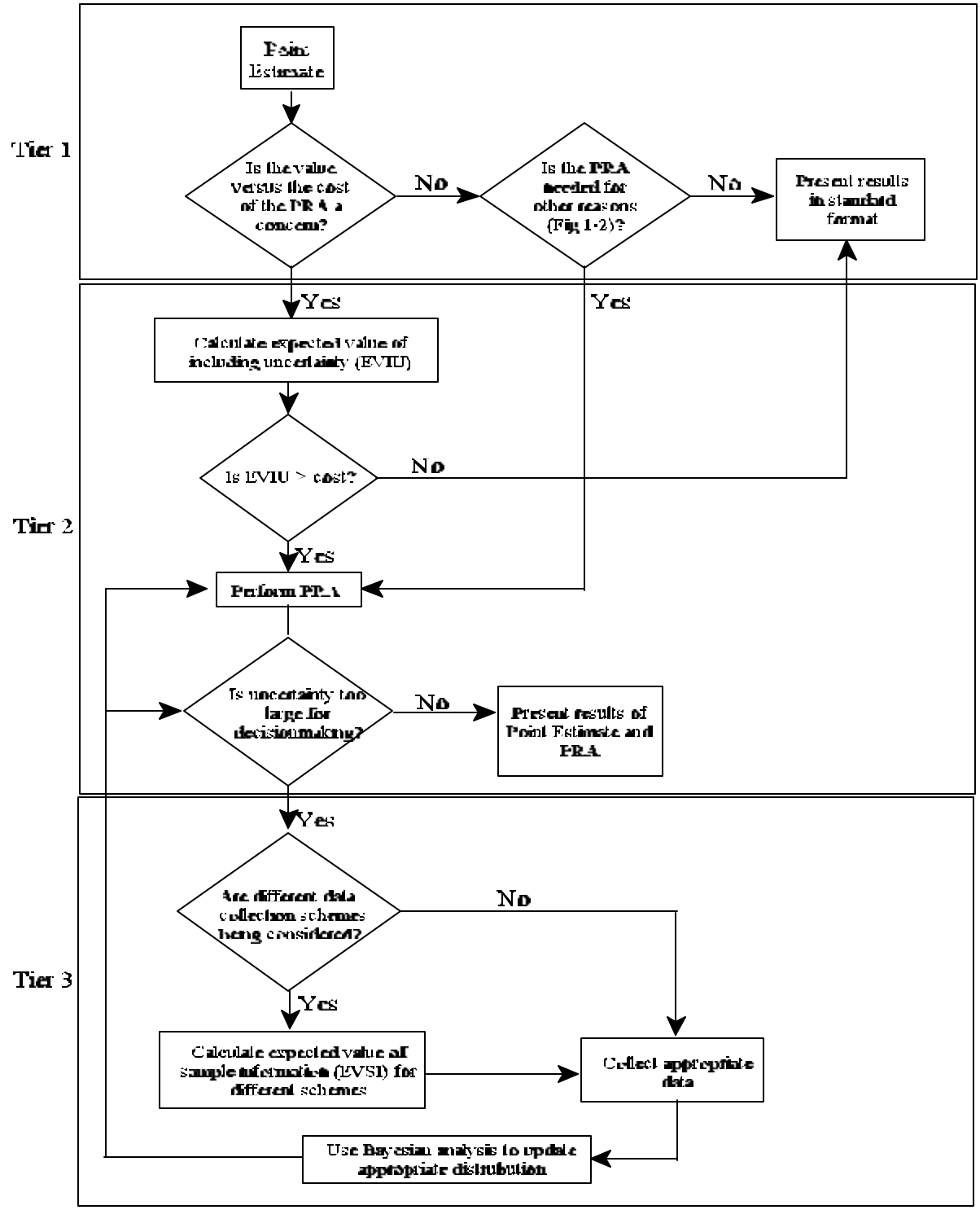


Figure E-9. Conceptual model for evaluating the expected value of including uncertainty in a Bayesian Monte Carlo Analysis.

In order to carry out VOI calculations, a loss function must be assumed. Definition of the loss function may be complex due to multiple decision goals and/or multiple decision makers and may be difficult to capture in an equation. Finally, for Bayesian analysis and the calculation of the EVSI to be helpful, one or more sources of new data must exist. In addition, some information must be available about this data since a likelihood function describing its probability distribution must be assumed.

REFERENCES FOR APPENDIX E

Brattin, W.J., T.M. Barry and N. Chiu. 1996. Monte Carlo modeling with uncertainty probability density functions. *Human Eco. Risk Assess*. 2(4): 820-840.

Buck, R.J., K.A. Hammerstrom and P.B. Ryan. 1997. Bias in population estimates of long-term exposure from short-term measurements of individual exposure. *Risk Anal.* 17: 455-466.

Burmaster, D.E. and K.M. Thompson. 1997. Estimating exposure point concentrations for surface soils for use in deterministic and probabilistic risk assessments. *Human Eco. Risk Assess.* 3(3): 363-384.

Clemen, R.T. 1996. *Making Hard Decisions: An Introduction to Decision Analysis.* Duxbury Press, Pacific Grove, CA.

Cliff, A. and J. Ord. 1981. *Spatial Processes*. Pion. London.

Clifford, P.A., Barchers, D.E., Ludwig, D.F., Sielken, R.L., Klingensmith, J.S., Graham, R.V., and Banton, M.I. 1995. An approach to quantifying spatial components of exposure for ecological risk assessment. *Environ. Tox. Chem*. 14(5): 895-906.

Cressie, N. 1991. *Statistics for Spatial Data.* Wiley. New York, NY.

Dakins, M.E., J. E. Toll and M. J. Small. 1994. Risk-based environmental remediation: decision framework and role of uncertainty. *Environ. Tox. Chem* 13:1907-1915.

Dakins, M.E., J.E. Toll, M.J. Small, and K. P. Brand. 1996. Risk-based environmental remediation: Bayesian Monte Carlo Analysis and the expected value of sample information. *Risk Anal.* 16:67-79.

Dakins, M.E. 1999. The Value of the Value of Information. Human and Ecological Risk Assessment: 5(2):281-289.

Finkel, A.M. and J.S. Evans. 1987. Evaluating the benefits of uncertainty reduction in environmental health risk management. *J. of the Air Pollution Control Assoc.* 37:1164-1171.

Griffith, D.A. 1993. *Spatial Regression Analysis on the PC: Spatial Statistics Using SAS*. Association of American Geographers. Washington, DC.

Hora, S.C. 1992. Acquisition of expert judgment: examples from risk assessment. *J. of Energy Engin.* 118: 136-148.

Howson, C. and P. Urbach. 1989. *Scientific Reasoning: The Bayesian Approach*. Open Court. LaSalle, IL.

Isaaks, E. and R. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press. Oxford.

Morgan, G.M. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press. New York.

Ott, W.R. 1995. *Environmental Statistics and Data Analysis.* CRC Press. Boca Raton.

Patwardhan, A. and M.J. Small. 1992. Bayesian methods for model uncertainty analysis with application to future sea level rise. *Risk Anal.* 12:513-523.

Price, P.S., C.L. Curry, P. E. Goodrum, M.N. Gray, J. I. McCrodden, N.H. Harrington, H. Carlson-Lynch and R. E. Keenan. 1996. Monte Carlo modeling of time-dependent exposures using a microexposure event approach. *Risk Anal.* 16: 339-348.

Sivia, D.S. 1996. *Data Analysis: A Bayesian Tutorial*. Clarendon Press. Oxford.

Slob, W. 1996. A comparison of two statistical approaches to estimate long-term exposure distributions from short-term measurements. *Risk Anal.* 16: 195-200.

U.S. EPA. 1982. *Air Quality Criteria for Particulate Matter and Sulfur Oxides*. ECAO, Office of Research and Development. EPA/600/8-82/029.

U.S. EPA. 1997. *Exposure Factors Handbook. Update to Exposure Factors Handbook.* Office of Research and Development, NCEA. EPA/600/8-89/043, May 1989. August 1996.

Wallace, L.A., N, Duan and R. Ziegenfus. 1994. Can long-term exposure distributions be predicted from short-term measurements? *Risk Anal.* 14: 75-85.